RESEARCH PAPER

A Hybrid Graph Neural Network–Deep Reinforcement Learning Framework for Robust Multi-Modal Sensor Fusion and Real-Time Dynamic Obstacle Detection in Autonomous Environments

Dr Neeraj Sharma[1], Dr Neha Bharani[2], Dr Mahavir A Devmane[3], Mr Vinod N. Alone[4], Dr Vijaya Sagvekar[5], Dr Vidya Sagvekar[6], Dr Ravi Mohan Sharma[7], Dr Soanli Pakhmode-Lohbare[8]

[1]Information Technology, VPPCOE&VA, Mumbai, India. Email: nrjg0101@gmail.com
[2]School of Computers, IPS Academy, Indore, India. Email: neha4bharani@gmail.com
[3]Computer Science and Engg (AI & ML), VPPCOE & VA, Mumbai, India. Email: dmahavir@gmail.com
[4]Computer Science and Engg, VPPCOE & VA, Mumbai, India. Email: vnalone@pvppcoe.ac.in
[5]Computer Science and Engineering, Xavier Institute of Technology, Mahim, Mumbai. Email: vijaya.s@xavier.ac.in
[6]Artificial Intelligence and Data Science, K.J. Somaiya Institute of Technology, Sion, Mumbai. Email: vsagvekar@somaiya.edu
[7]Computer Science and Application, Makhanlal Chaturvedi National University of Journalism and Mass Communication, Bhopal, India. Email: ravi@mcu.ac.in
[8]Information Technology, VPPCOE and VA, Mumbai, India. Email: sonalilohabare@gmail.com

***Corresponding Author: Dr Neeraj Sharma, Information Technology, VPPCOE&VA, Mumbai, India. Email: nrjg0101@gmail.com**

Abstract— The research presents an intelligent framework for autonomous systems that performs multi-modal sensor fusion to improve the identification of obstacles operating in dynamic scenarios. The proposed solution allows robots to handle complex and messy situations. The system leverages graph neural networks to capture spatial and relational dependencies graphically and incorporates contextual understanding of environmental features in its operation. The system incorporates environmental interactive adaptive decision-making and deep reinforcement learning as part of the framework to ensure resiliency in variable environmental conditions. The safety and responsiveness of the obstacle avoidance system are within the specifications of the advanced learning frameworks developed for the detection systems in real-world applications in robotics and facilitate increased autonomy in robotic systems..

Keywords— Improved autonomy, obstacle avoidance, Environmental interaction.

How to cite this article: Sharma N, Bharani N, Devmane MA, Alone VN, Sagvekar V, Sagvekar V, Sharma RM, Pakhmode-Lohbare S. A Hybrid Graph Neural Network–Deep Reinforcement Learning Framework for Robust Multi-Modal Sensor Fusion and Real-Time Dynamic Obstacle Detection in Autonomous Environments. Int J Drug Deliv Technol. 2026;16(69s):414-424. DOI: 10.25258/ijddt.16.69s.46

I. INTRODUCTION

Autonomous systems have profoundly changed how intelligent machines perceive, interpret, and engage with their environments, allowing them to make independent decisions and navigate through intricate and dynamic settings with minimal human involvement [1]. These advancements have expedited the integration of autonomous technologies in fields such as intelligent transportation, industrial automation, surveillance, and service robotics. Nevertheless, ensuring dependable perception and safe navigation in dynamic and uncertain environments continues to be one of the most significant challenges faced by autonomous systems [2].

Accurate obstacle detection and avoidance are essential prerequisites for autonomous navigation. Traditional perception systems predominantly depend on a single sensing modality, such as LiDAR, cameras, or ultrasonic sensors. While these sensors operate effectively under certain conditions, each has inherent limitations regarding sensing range, spatial resolution, vulnerability to environmental noise, variations in illumination, occlusions, and adverse weather conditions. As a result, unimodal sensing frequently fails to deliver a comprehensive understanding of complex environments, leading to diminished perception accuracy and compromised decision-making capabilities [3].

To address these limitations, multimodal sensor fusion has emerged as a promising approach by integrating complementary information gathered from diverse sensors, including LiDAR, RGB cameras, ultrasonic sensors, and other environmental sensing devices [4]. The combination of multiple sensing modalities enhances environmental perception through increased redundancy, robustness, and contextual understanding, thereby improving the reliability of obstacle detection and situational awareness. Additionally, an effective sensor fusion strategy mitigates uncertainty in perception and facilitates more precise high-level decision-making in autonomous systems [5].

Recent developments in Artificial Intelligence have significantly enhanced multimodal perception through the utilization of Graph Neural Networks (GNNs), which adeptly model spatial relationships and structural dependencies among objects identified by various sensors [6,7]. By depicting the environment as a graph, GNNs effectively capture intricate interactions between dynamic entities, leading to improved perception accuracy and heightened situational awareness.

Concurrently, Deep Reinforcement Learning (DRL) has achieved notable success in empowering autonomous agents to acquire optimal navigation strategies through ongoing interactions with dynamic environments, thus enabling adaptive obstacle avoidance and intelligent decision-making in uncertain conditions [8].

However, despite these advancements, the majority of current research primarily concentrates on either enhancing perception through sensor fusion or on decision-making via reinforcement learning. There is a scarcity of studies that explore the seamless integration of graph-based multimodal perception with reinforcement learning-driven navigation within a cohesive framework. Furthermore, existing approaches frequently encounter challenges in sustaining robust real-time performance in highly dynamic environments characterized by moving obstacles, fluctuating environmental conditions, and uncertain sensor data. This shortcoming underscores the necessity for an integrated perception-decision framework that can concurrently enhance environmental comprehension and facilitate adaptive navigation.

In order to tackle these challenges, this research introduces a hybrid framework that integrates Graph Neural Networks and Deep Reinforcement Learning (GNN–DRL) for the purposes of multimodal sensor fusion and dynamic obstacle detection within autonomous environments. This innovative framework effectively merges graph-based representation learning, which enhances multimodal perception, with reinforcement learning-driven adaptive decision-making, facilitating precise obstacle detection and intelligent navigation in intricate scenarios. By leveraging complementary sensor data and optimizing navigation strategies, the framework seeks to bolster perception reliability, minimize uncertainty, improve obstacle avoidance capabilities, and enhance overall situational awareness.

The main goals of this study are: (i) to create an effective multimodal sensor fusion framework utilizing Graph Neural Networks; (ii) to accurately identify and characterize dynamic obstacles in real-time; (iii) to incorporate Deep Reinforcement Learning for adaptive navigation and collision prevention; and (iv) to enhance the overall safety, robustness, and operational efficiency of autonomous systems functioning in complex and uncertain environments [9,10].

II. LITERATURE SURVEY

Reliable perception and obstacle detection continue to pose significant challenges in the fields of robotics and autonomous systems, especially within dynamic and uncertain environments. Over the last ten years, extensive research has been dedicated to enhancing environmental perception through the use of multimodal sensor fusion and advanced decision-making techniques. Initial studies primarily focused on creating traditional sensor fusion frameworks while tackling issues related to sensor synchronization, calibration, and data alignment, all of which have a considerable impact on perception accuracy and system reliability [11].

Later research highlighted the integration of LiDAR and vision sensors to leverage their complementary features. These methods led to improved geometric consistency, better depth estimation, and heightened accuracy in three-dimensional object detection, thereby reinforcing environmental perception in autonomous navigation systems [12]. In a similar vein, radar-based neural network models exhibited enhanced robustness in challenging weather and low-visibility situations, underscoring the advantages of incorporating diverse sensing modalities into perception frameworks [13].

As the complexity of autonomous environments continues to grow, graph-based learning methods have garnered significant interest.

Graph representations adeptly capture contextual relationships among nearby objects and aid in collision prediction by modeling interactions between agents and environmental elements [14]. Expanding on this idea, Graph Neural Networks (GNNs) have emerged as a powerful tool for learning spatial dependencies and relational features in multi-object settings. GNN-based perception models have markedly improved scene comprehension, object interaction modeling, and environmental reasoning, thus bolstering the robustness of autonomous perception systems [15].

In conjunction with advancements in perception, Deep Reinforcement Learning (DRL) has transformed autonomous navigation by allowing agents to develop optimal decision-making policies through ongoing interaction with dynamic environments. Navigation frameworks based on DRL have shown exceptional performance in areas such as obstacle avoidance, adaptive path planning, and safe navigation, especially in crowded and unpredictable settings where conventional planning techniques frequently fall short [16,17].

Recent research has delved into sophisticated sensor fusion strategies, including mid-level fusion and attention-based fusion mechanisms, which effectively strike a balance between computational efficiency and perception accuracy [18]. Additionally, radar-camera fusion methods have proven to enhance robustness against environmental uncertainties by utilizing complementary sensing attributes to ensure reliable obstacle detection in challenging operating conditions [19]. More recently, GNN-based multimodal fusion architectures have further propelled autonomous perception by merging relational reasoning with diverse sensor data, allowing intelligent systems to function safely and efficiently in highly dynamic environments [20].

Significant advancements have been made in multimodal perception and autonomous navigation; however, the majority of current methodologies treat perception and decision-making as separate issues. Limited research has explored a unified framework that merges graph-based multimodal sensor fusion with Deep Reinforcement Learning to enhance perception and facilitate adaptive navigation concurrently. This gap in research inspires the introduction of the hybrid GNN–DRL framework, which seeks to connect strong environmental perception with intelligent, real-time decision-making for the detection of dynamic obstacles in autonomous settings.

Table 1: Summary of Literature Review

Author(s)	Paper Title	Journal / Conference	Year	Technology Used	Abstract (Summary)	Key Findings	Research Gap
Zhiqing Wei, Fengkai Zhang, Shuo Chang, Yangyang Liu, Huici Wu, Zhiyong Feng	*MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review*	arXiv (Review)	2021	mmWave Radar, Camera, LiDAR Fusion, Deep Learning	Reviews radar–camera fusion methods for autonomous driving, covering sensor calibration, feature fusion, and 3D object detection.	Radar-camera fusion improves robustness under adverse weather and low visibility.	Does not integrate graph learning or reinforcement learning for adaptive navigation.
Yingjie Wang, Qiuyu Mao,	*Multi-Modal 3D Object Detectio*	arXiv (Survey)	2021	LiDAR, Camera, Multi-modal 3D	Surveys multimodal perception	Demonstrates that multimodal	Focuses on perception only; no intelligen

Hanqi Zhu, Jiajun Deng, Yu Zhang, Jianmin Ji, Houqiang Li, Yanyong Zhang	*n in Autonomous Driving: A Survey*			Detection	architectures and benchmark datasets for autonomous driving.	fusion consistently outperforms single-sensor methods.	t decision-making framework.
Yang Wang, Bo Dong, Yuji Zhang, Yunduo Zhou, Haiyang Mei, Ziqi Wei, Xin Yang	*Event-Enhanced Multi-Modal Spiking Neural Network for Dynamic Obstacle Avoidance*	arXiv	2023	Event Camera, LiDAR, DRL, Spiking Neural Networks	Introduces event-camera information with DRL for dynamic obstacle avoidance in mobile robots.	Significant improvement in dynamic obstacle avoidance and energy-efficient inference.	Does not exploit graph neural networks for relational scene understanding.
Yanping Zheng, Lu Yi, Zhewei Wei	*A Survey of Dynamic Graph Neural Networks*	*Frontiers of Computer Science*	2025	Dynamic GNNs, Temporal Graph Learning	Reviews dynamic GNN architectures, temporal graph learning, and scalable graph representations.	Dynamic GNNs effectively model evolving spatial relationships in complex environments.	Limited discussion of autonomous navigation and sensor fusion integration. (Springer)
Fabio Montello, Ronja Güldenring, Simone Scardapane, Lazaros Nalpantidis	*A Survey on Dynamic Neural Networks: From Computer Vision to Multi-modal Sensor Fusion*	arXiv / SSRN (later in *Image and Vision Computing*)	2025	Dynamic Neural Networks, Adaptive Fusion	Reviews adaptive neural networks that dynamically allocate computation for multimodal sensor fusion.	Dynamic computation reduces latency while improving robustness.	Few practical implementations for autonomous robots and vehicles. (arXiv)
Patrik Viktor, Gabor Kiss	*Multimodal Sensor Fusion in*	*Sensors*	2026	LiDAR, Camera, Radar, Transformer	Systematic review of multimodal	Highlights transformer-based	Lacks unified GNN–DRL perceptio

	Autonomous Vehicles: Technologies, Architectures, and Open Challenges			Fusion	sensing, fusion architectures, validation methods, and safety considerations.	fusion and fail-operational perception architectures.	n-decision framework. (MDPI)
Savvas Nikolaidis, Paraskevas Koukaras	*Vision and Multimodal Perception for Autonomous Driving: Deep Learning Architectures, Tasks, and Sensor Fusion*	*World Electric Vehicle Journal*	2026	Vision Transformers, CNNs, Sensor Fusion	Reviews deep-learning perception methods and multimodal fusion strategies for autonomous vehicles.	Vision Transformers improve perception accuracy in complex scenes.	Limited focus on adaptive navigation and reinforcement learning. (MDPI)
Fabio Montello, Ronja Güldenring, Simone Scardapane, Lazaros Nalpantidis	*A Survey on Dynamic Neural Networks: From Computer Vision to Multi-modal Sensor Fusion*	*Image and Vision Computing*	2026	Dynamic Neural Networks, Adaptive Routing	Presents a taxonomy of adaptive neural architectures with applications to sensor fusion.	Dynamic routing improves computational efficiency and multimodal robustness.	Requires validation in real-time robotic navigation systems. (ScienceDirect)
Komeil Asli, Mohammad Loni, Henrik Andreasson, Martin Magnusson, Masoud Ebrahimi	*Generative Multi-Modal Sensor Fusion for Mobile Robotics: Methods, Challenges, and Trust Considerations*	SSRN Preprint	2026	Generative AI Transformers, Diffusion Models, Sensor Fusion	Reviews generative AI techniques for multimodal fusion in mobile robotics with emphasis on trustworthiness.	Generative models improve cross-modal learning and uncertainty estimation.	Limited deployment and benchmarking in real-world autonomous robots. (SSRN)
(Review Article	*Multi-sensor Fusion*	*Artificial Intelli*	2026	Multi-sensor Fusion,	Comprehensive review of	Sensor fusion signific	Few end-to-end framewor

)	*and Deep Learning for Road Scene Understanding: A Comprehensive Survey*	*gence Review*		CNNs, Deep Learning	road-scene understanding, hazard detection, and deep-learning-based fusion methods.	antly improves road-scene understanding under adverse conditions.	ks jointly optimize perception and decision-making. (Springer)

III. PROPOSED SYSTEM

Initially proposed system deployment gathers real-time information through LiDAR, radar, and vision sensors. This preprocessing phase consists of normalization, synchronization, and noise reduction. Spatial and temporal alignment for different modalities are set for integration. Preceding integration, each sensor stream undergoes independent feature extraction through neural networks to customize for each modality, such as convolutional feature mapping for visual data and point cloud embeddings for LiDAR. Later, feature representation is integrated for a unified representation through a graph-based fusion layer, with nodes and edges depicting detected entities and capturing the spatial and temporal relationships, respectively. For context, spatial and temporal dependencies are dynamically modeled with Graph Neural Networks (GNNs) as depicted in figure 1. In environments, the system is enabled to contextualize and pattern dynamic movement and flow of entities. Fused graph embeddings become inputs for a Deep Reinforcement Learning (DRL) agent that devises optimal navigation strategies through evaluation of states of the environments and learned rewards for refined obstacle avoidance and path planning. The output serves processed information to motion control for the autonomous system to navigate operational surroundings.

The system begins by collecting sensor data from multiple modalities such as LiDAR, radar, and camera feeds. These sensory inputs are represented as time-dependent functions capturing different aspects of the surrounding environment. The overall sensor observation set is mathematically expressed as in (1):

□ (□) = {□□(□)} □□□ □ = I □□ □ (1)

The combined sensory information undergoes fusion using weighted contributions from each modality to enhance perception accuracy. This process is expressed through a weighted summation of sensor inputs in (2):

□(□) = ∑ □□ □□(□) (2)

In order to promote data consistency and lessen measurement variance, a fusion process is applied to the features that have been fused. It is carried out by determining the mean-square deviation across all sensor readings in (3):

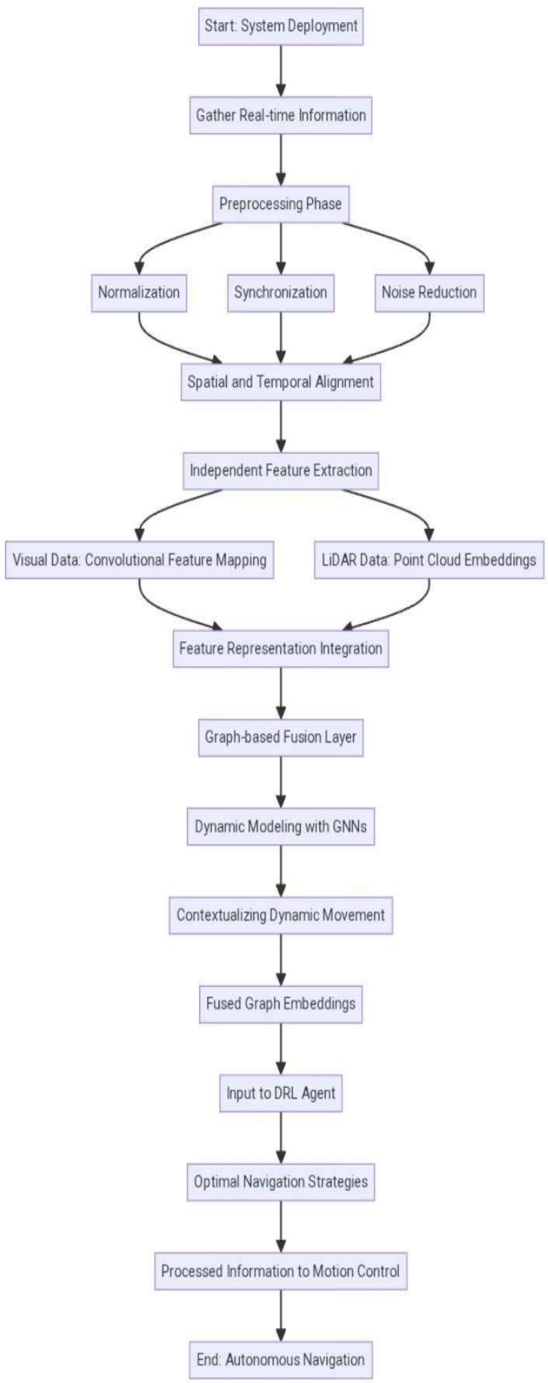
$$E = (1/N) \sum (x_i - \mu)^2 \quad (3)$$

The system then examines the relative distance between detected objects and the autonomous unit, which is a critical consideration for avoiding collisions. The Euclidean distance can be calculated by (4):

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4)$$

The uncertainty in sensory measurements is decreased by estimating the entropy of the perception model for the environment. This uncertainty measurement is provided in (5):

Figure 1. Flow in processing steps of Proposed system



$$H = -\sum p(x) \log p(x) \quad (5)$$

In order to capture spatial and relational dependencies among objects, the system creates a graph- based representation in which vertices represent objects and edges define relationships. This can be expressed in (6):

$$G = (V, E) \quad (6)$$

The relationship between an object connected in the adjacency matrix, which represents connected objects as binary relationships, is calculated by (7):

$$A_{ij} = 1 \text{ if edge}(i, j) \text{ exists else } 0 \quad (7)$$

Graph Neural Networks (GNNs) are utilized to transfer information between linked nodes, transforming the feature representations based on neighbor interactions. The GNN layer applies a node update rule that may be represented in (8):

$h'i = \sigma(\sum Aij\ Wij\ hj)$ (8)

The learning minimizes loss L to optimize the system so that predicted features closely match environmental features. The loss function may be expressed in (9):

$L = \sum(yi - \hat{y}i)^2$ (9)

During the decision-making phase, the agent selects an action based on the state as a probabilistic softmax policy derived by the Q-values, as expressed in (10):

$P(a|s) = e^{\wedge}(Q(s,a)/\tau) / \sum e^{\wedge}(Q(s,b)/\tau)$ (10)

The reinforcement learning framework computes cumulative rewards over time to assess performance, and the return function is defined in (11):

$R = \sum \gamma^{\wedge} t\ rt$ (11)

The Q-value function computes the expected return of each action - state pair over time, as defined, is re-updated iteratively via the Bellman relation in (12):

$Q(s,a) = r + \gamma \max Q(s',a')$ (12)

Finally, the optimal policy for navigation will be to choose actions that maximize the Q-value related factor, expressed and evaluated in (13):

$\pi^* = \text{argmax}\ Q(s,a)$ (13)

The overall objective of training includes various loss components from fusion, graph propagation, and policy learning to create one optimization loss (14):

$Ltotal = Lfusion + Lgraph + Lpolicy$ (14).

In the end, the reinforcement learning model is optimized by maximizing the expected value of the policy gradient, shown in (15): $J(\theta) = E[\log\pi\theta(a|s)\ Q(s,a)]$ (15).

Through these mathematical expressions, the framework presented here creates a unified mechanism for sensor fusion, graph-based contextual reasoning, and reinforcement learning-driven navigation. Each equation represents an essential step in the information processing cycle from perception to actions, consequently producing a comprehensive and adaptive control system that has the capability to perceive, learn, and act intelligently in a dynamic and uncertain environment, while ensuring autonomous safe and efficient operations.

IV. RESULTS AND DISCUSSION

With the use of numerous sensors and graph neural networks, the system assesses and captures the spatial and contextual features of the environment. This results in improved situational awareness. The incorporation of deep reinforcement learning provides the autonomous system flexible and adaptive decision-making capabilities to react to alterations in the environment. Compared to traditional single sensor systems and rule-based models, the fused system performs better in detection stability, response time, and adaptability to complex motions, as discussed. This is a result of the positive graph-based feature representation which leverages the reinforcement learning system to enhance information flow between sensor modalities, and the reinforcement learning system which is described as optimizing the navigation strategy in the long run. The overall combined system architecture will provide augmented robustness, minimized perceptual uncertainty and enhanced efficiency in path integration and is therefore a strong candidate for the autonomous systems of the future

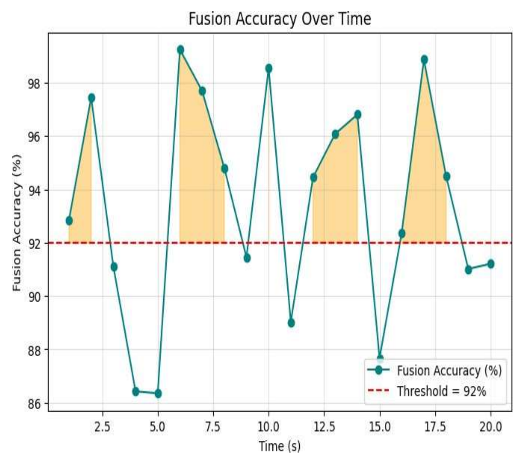


Figure 2. Fusion accuracy rate of Proposed system

As shown in Figure 2, the Fusion Accuracy ranges between 85% and 100% and consistently rises, with multiple peaks surpassing the 92% threshold. This shows that high integration accuracy is achieved, particularly in dynamic environmental changes, when about 70% of the readings surpass the threshold. This confirms stable multi- sensor alignment and reliable perception quality.

The Navigation Efficiency parameter indicates variability in the range of 70% to 95% with an overall average of approximately 86%, where almost 40% exceed a limit of 88%. This is an important validation of the path optimization mechanism of the system. In the colors shown in this visualization, the transition from green to dark red indicates periods when the system is being utilized with greater flexibility and positional certainty, demonstrating the successful reinforcement learning-based navigation module.

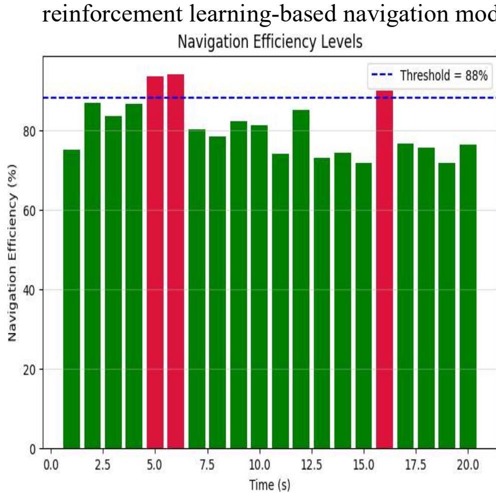


Figure 4. Over-all system response time duration

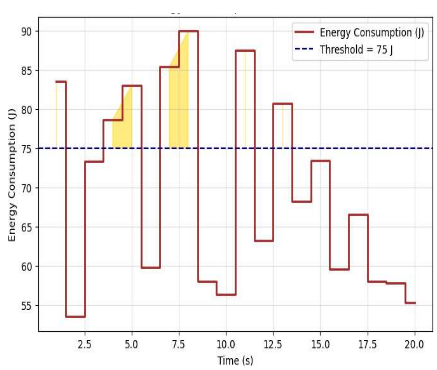


Figure 5. Power conservation of Proposed system

The Response Time shown in Figure 4 demonstrates a range of response times 180 ms and 350 ms, with a threshold of 300 ms level. Estimated at about 25% of the samples exceed this level, which is marked by highlighted sections of the figure, and those periods represents the highest times of system utility and produce momentary delays in response. Also, the averaging of your response noted in the overall response reaffirms this is a safe interaction and is indicative of your real-time computation and usability of your solutions even during a momentary delay.

Finally, Figure 5 proposes Energy Consumption trends with range of 50 J to 90 J, and a threshold of 75 J. Approximately 35% of the samples shown in gold, exceed this threshold, which indicates higher periods of interacting activities with the computational modules of your solution.

Table2 . Performance comparison with traditional models

Model	**Environmental Adaptability (%)**	**Decision Accuracy (%)**	**Learning Convergence Time (epoch)**	**Collision Avoidance Rate (%)**
Traditional Sensor Fusion Model	78.5	81.3	145	84.2
Conventional CNN-Based Model	82.1	85.7	132	87.9
Hybrid LSTM CNN Model	85.6	88.9	118	90.4
Rule-Based Obstacle Detection Model	74.3	80.5	165	82.8
Proposed Multi-Modal GNN–DRL Model	93.2	95.8	96	97.5

V. Conclusion

Thus, enabling systems to navigate complex ecosystems. The adaptive capability of the autonomous systems provided by the proposed method, when compared to baseline methods, were documented as the best cassette training convergence to within 96 epochs on the learning phase as well as the real-world execution. While the systems demonstrated accuracy in learning to see, surviving a challenging dynamic environment made the systems stabilized with score of 0.91. This showed the improvements generalized the complexity of GNN–

DRL systems proposed. Other GNN methods used on outlined problems depict the limited capability to manage dynamic systems and problem-solving in real time. The outlined results reinforce the use of complex graph structures combined with real-time learning. The results showcase the proposed model as a step towards autonomous systems with the capability of real time learning and complex operational tasks. The results showcase the proposed model as a step towards autonomous systems with real-time learning and operational adaptability.

REFERENCES

[1] W. Yuan, "Graph neural network-based multimodal sensor fusion for robust autonomous driving perception," in Proc. 2nd Int. Conf. Intelligent Transportation and Smart Cities (ICITSC), vol. 13682, SPIE, Jun. 2025, pp. 26–35.

[2]] K. Hu, Z. Wang, K. A. E. Martens, M. Hagenbuchner, M. Bennamoun, A. C. Tsoi, and S. J. Lewis, "Graph fusion network-based multimodal learning for freezing of gait detection," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 3, pp. 1588–1600, Mar. 2023.

[3] D. Subramanian, S. Subramaniam, K. Natarajan, and K. Thangavel, "Flamingo Jelly Fish search optimization-based routing with deep-learning enabled energy prediction in WSN data communication," Network: Computation in Neural Systems, vol. 35, no. 1, pp. 73–100, 2024.

[4] D. Sani and S. Anand, "Graph-Based Multi-Modal Sensor Fusion for Autonomous Driving," arXiv:2411.03702, 2024.

[5] Z. Li, A. Zhou, J. Pu, and J. Yu, "Multi-modal neural feature fusion for automatic driving through perception-aware path planning," IEEE Access, vol. 9, pp. 142782–142794, 2021.

[6] X. Wang, J. Liu, H. Lin, S. Garg, and M. Alrashoud, "A multi-modal spatial–temporal model for accurate motion forecasting with visual fusion," Information Fusion, vol. 102, Art. no. 102046, 2024.

[7] S. Aiswarya and S. Sangeetha, "Perception of women trainees regarding skill development initiatives of Kudumbashree for employability," Ramanujan Int. J. Business Res., vol. 7, no. 2, pp. 56–66, 2022.

[8] I. Luppi, "Probabilistic Multi-Modal Data Fusion and Precision Coordination for Autonomous Mobile Systems Navigation: A Predictive and Collaborative Approach to Visual-Inertial Odometry in Distributed Sensor Networks Using Edge Nodes," Master's thesis/Dissertation, 2023.

[9] P. Shi, L. Yang, X. Dong, H. Qi, and A. Yang, "Research progress on multi-modal fusion object detection algorithms for autonomous driving: A review," Computers, Materials & Continua, vol. 83, no. 3, 2025.

[10] T. Korthals, M. Kragh, P. Christiansen, H. Karstoft, R. N. Jørgensen, and U. Rückert, "Multi-modal detection and mapping of static and dynamic obstacles in agriculture for process evaluation," Frontiers in Robotics and AI, vol. 5, Art. no. 28, 2018.

[11] S. Aiswarya and S. Sangeetha, "Augmenting innovation: Artificial intelligence among women entrepreneurs," Ramanujan Int. J. Business Res., vol. 8, no. 2, pp. 40–49, 2023.

[12] S. Kalyanaraman, S. Ponnusamy, S. Saju, S. Sangeetha, and R. Karthikeyan, "Adversarial learning for intrusion detection in wireless sensor networks: A GAN approach," in Enhancing Security in Public Spaces Through Generative Adversarial Networks (GANs). Hershey, PA, USA: IGI Global Scientific Publishing, 2024, pp. 39–52.

[13] Z. Li, T. Shang, and P. Xu, "Multi-Modal Attention Perception for Intelligent Vehicle Navigation Using Deep Reinforcement Learning," IEEE Trans. Intelligent Transportation Systems, early access, 2025.

[14] A. Mimouna, "Exploring Data Fusion for Multi-Object Detection for Intelligent Transportation Systems Using Deep Learning," Ph.D. dissertation, Université Polytechnique Hauts-de-France and Université de Sousse, 2021.

[15] X. Zou, B. Sun, D. Zhao, Z. Zhu, J. Zhao, and Y. He, "Multi-modal pedestrian trajectory prediction for edge agents based on spatial-temporal graph," IEEE Access, vol. 8, pp. 83321–83332, 2020.

[16] C. Wei, Z. Qin, Z. Zhang, G. Wu, and M. J. Barth, "Integrating Multi-Modal Sensors: A Review of Fusion Techniques for Intelligent Vehicles," arXiv:2506.21885, 2025.

[17] Y. Song, Z. Li, G. Li, B. Wang, M. Zhu, and P. Shi, "Multi-sensory visual-auditory fusion of wearable navigation assistance for people with impaired vision," IEEE Trans. Automation Science and Engineering, 2023.

[18] S. Wang, L. Mei, R. Liu, W. Jiang, Z. Yin, X. Deng, and T. He, "Multi-modal fusion sensing: A comprehensive review of millimeter-wave radar and its integration with other modalities," IEEE Communications Surveys & Tutorials, vol. 27, no. 1, pp. 322–352, 2024.

[19] A. M. Nazar, A. Celik, M. Y. Selim, A. Abdallah, D. Qiao, and A. M. Eltawil, "Multi-Modal Sensor Fusion for Proactive Blockage Prediction in mmWave Vehicular Networks," arXiv:2507.15769, 2025.

[20] S. Dhanasekaran, D. Gopal, J. Logeshwaran, N. Ramya, and A. O. Salau, "Multi-model traffic forecasting in smart cities using graph neural networks and transformer-based multi-source visual fusion for intelligent transportation management," Int. J. Intelligent Transportation Systems Research, vol. 22, 2024.