A Hybrid AI Framework for Autonomous Driving Across Diverse Traffic and Weather Conditions

Navin Kumar Sehgal[1]

[1]Tech Department, IITM College of Engineering Bahadurgarh, India

Publication Date: 2026/07/29

Abstract: Autonomous driving systems must operate safely and reliably under diverse traffic densities, varying weather conditions, and dynamic road environments while maintaining real-time performance and low computational cost. However, many existing autonomous driving frameworks rely on expensive LiDAR sensors, high-performance computing hardware, and cloud-based processing, limiting their practical deployment in cost-sensitive applications. This paper proposes a Hybrid AI Framework for Autonomous Driving Across Diverse Traffic and Weather Conditions that integrates Segment Anything Model 2 (SAM2) for semantic scene segmentation, transformer-based multimodal sensor fusion for comprehensive environmental perception, reinforcement learning (RL) for adaptive decision-making, and a neural vehicle controller for continuous steering, throttle, and brake control. To reduce system cost and computational complexity, the framework employs a low-cost sensor suite comprising an RGB camera, automotive radar, ultrasonic sensors, GPS, accelerometer, gyroscope, odometer, magnetometer, temperature sensor, humidity sensor, and vibration sensor, eliminating the need for expensive LiDAR systems. The entire architecture is optimized for deployment on the low-cost NVIDIA Jetson embedded edge-computing platform, enabling real-time processing with reduced latency, lower power consumption, and improved operational efficiency. The proposed framework is designed to handle challenging edge-case scenarios, including sudden obstacles, adverse weather, dense traffic, and low-visibility conditions, by dynamically adapting sensor fusion and driving policies. Experimental evaluation demonstrates that the proposed approach achieves over 90% perception and decision-making accuracy, while maintaining stable vehicle control, smooth steering, adaptive throttle regulation, and timely braking across varying traffic and weather conditions. The results further indicate significant improvements in driving safety, vehicle stability, collision avoidance, and computational efficiency compared with conventional autonomous driving approaches. The proposed hybrid architecture provides a scalable, cost-effective, and practical solution for next-generation intelligent autonomous vehicles operating in real-world environments.

Keywords: *Autonomous Driving, Artificial Intelligence, Segment Anything Model 2 (SAM2), Transformer-Based Sensor Fusion, Reinforcement Learning, Edge Computing, Semantic Segmentation, Vehicle Control, Intelligent Transportation Systems*.

How to Cite: Navin Kumar Sehgal (2026) A Hybrid AI Framework for Autonomous Driving Across Diverse Traffic and Weather Conditions. *International Journal of Innovative Science and Research Technology*, 11(7), 1621-1632. https://doi.org/10.38124/ijisrt/26jul923

I. INTRODUCTION

Autonomous driving has emerged as one of the most significant advancements in artificial intelligence (AI), computer vision, robotics, and intelligent transportation systems. By enabling vehicles to perceive their surroundings, interpret dynamic road environments, make intelligent decisions, and execute appropriate control actions without continuous human intervention, autonomous vehicles have the potential to improve road safety, reduce traffic congestion, enhance transportation efficiency, and provide greater mobility for people. Recent developments in deep learning, multimodal sensor fusion, reinforcement learning, and embedded computing have accelerated the development of intelligent driving systems capable of operating in increasingly complex environments. However, achieving reliable and safe autonomous driving under diverse traffic conditions, adverse weather, and unexpected road scenarios remains a major research challenge.

Most existing autonomous driving systems employ multiple high-cost sensors such as LiDAR, high-resolution cameras, and powerful Graphics Processing Units (GPUs) to achieve accurate environmental perception and vehicle control. Although these systems demonstrate excellent performance in controlled environments, their deployment is often limited by high computational requirements, increased hardware costs, and significant power consumption. Furthermore, many perception models experience degraded performance in challenging situations such as rain, fog, nighttime driving, dense urban traffic, temporary road obstructions, and other edge-case scenarios. These limitations highlight the need for autonomous driving frameworks that are computationally efficient, economically feasible, and capable of maintaining reliable performance under diverse environmental conditions.

Recent advances in foundation models have introduced new opportunities for improving autonomous vehicle perception. Segment Anything Model 2 (SAM2) provides generalized semantic segmentation capabilities that enable accurate identification of roads, lane markings, vehicles, pedestrians, traffic signs, and previously unseen objects without requiring extensive task-specific retraining. Similarly low computation transformer-based architectures can be used remarkable in multimodal sensor fusion by learning relationships among heterogeneous sensor modalities and predicting outcomes. Unlike conventional fusion methods that use fixed weighting strategies, transformers employ self-attention mechanisms to adaptively prioritize the most reliable sensor information according to changing environmental conditions. Reinforcement Learning (RL) has also shown considerable promise in autonomous driving by enabling vehicles to learn optimal driving policies through interaction with dynamic environments, resulting in adaptive steering, acceleration, and braking decisions that improve safety and driving efficiency.

Based on these conclusions, this paper proposes a Hybrid AI Autonomous Driving Framework that integrates SAM2-based semantic segmentation, transformer-based multimodal sensor fusion, reinforcement learning-based decision planning, and a neural vehicle controller into a unified perception-to-control architecture. To improve the practicality and affordability of autonomous driving, the proposed framework replaces expensive LiDAR-based perception with a cost-effective multimodal sensor suite consisting of low cost RGB camera, automotive radar, ultrasonic sensors, GPS, accelerometer, gyroscope, odometer, magnetometer, temperature sensor, humidity sensor, and vibration sensor. Furthermore, the entire framework is optimized for deployment on the NVIDIA Jetson embedded edge-computing platform, enabling real-time inference while reducing computational complexity, energy consumption, and communication latency associated with cloud-based processing.

The proposed framework is specifically designed to operate reliably under varying weather conditions, including sunny, cloudy, rainy, foggy, and nighttime environments, as well as different traffic densities ranging from low traffic to highly congested urban roads. By combining semantic scene vectors , sensor fusion, reinforcement learning, and neural vehicle control, the framework effectively addresses challenging almost all scenarios such as sudden obstacle appearance, reduced visibility, unexpected traffic behavior, and sensor uncertainty. Experimental evaluation demonstrates that the proposed approach achieves an overall perception and decision-making accuracy exceeding 90%, while maintaining smooth steering control, adaptive throttle regulation, timely braking responses, and high levels of safety, stability, and driving efficiency. The proposed Hybrid AI framework therefore provides a computationally efficient, cost-effective, and scalable solution for next-generation autonomous driving systems suitable for real-world deployment under diverse environmental and operational conditions.

II. LITERATURE REVIEW

➢ *Background*

Autonomous driving has emerged as one of the most active research areas in artificial intelligence, computer vision, robotics, and intelligent transportation systems. The primary objective of autonomous vehicles is to perceive their surroundings, understand complex road environments, make intelligent driving decisions, and safely control the vehicle with minimal or no human intervention. Achieving reliable autonomous driving requires seamless integration of perception, localization, sensor fusion, planning, decision-making, and vehicle control. Although considerable progress has been made in recent years, developing autonomous driving systems capable of operating safely under diverse traffic conditions, adverse weather, and unexpected edge-case scenarios remains a significant research challenge.

Early autonomous driving systems relied heavily on rule-based algorithms and classical computer vision techniques for lane detection, obstacle detection, and traffic sign recognition. These methods used handcrafted image features, edge detection, Hough transforms, and template matching to identify road structures and surrounding objects (Dickmanns, 2007; Pomerleau, 1989). While computationally efficient, these techniques suffered from poor robustness under varying illumination, weather changes, occlusions, and dynamic traffic environments. Consequently, their practical applicability in real-world autonomous driving was limited.

➢ *Current Studies*

The emergence of deep learning significantly transformed autonomous vehicle perception. Convolutional Neural Networks (CNNs) demonstrated remarkable success in image classification, object detection, semantic segmentation, and scene understanding (Krizhevsky et al., 2012). Subsequently, advanced object detection algorithms such as Faster R-CNN (Ren et al., 2015), YOLO (Redmon et al., 2016), and SSD (Liu et al., 2016) substantially improved the detection of vehicles, pedestrians, cyclists, and traffic signs. These approaches enabled autonomous vehicles to perform robust environmental perception in real-time while achieving high detection accuracy.

Semantic segmentation has become an essential component of autonomous driving perception systems. Fully Convolutional Networks (FCNs) (Long et al., 2015), U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), DeepLab (Chen et al., 2018), and PSPNet (Zhao et al., 2017) enabled pixel-level classification of driving scenes, allowing accurate identification of roads, lane markings, sidewalks, vehicles, pedestrians, vegetation, and obstacles.

Recently, foundation models have introduced a new direction in computer vision by providing generalized scene understanding without requiring task-specific retraining. Segment Anything Model (SAM) and its successor Segment Anything Model 2 (SAM2) have demonstrated exceptional

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

zero-shot segmentation capabilities across diverse visual environments (Kirillov et al., 2023; Ravi et al., 2024). Unlike conventional segmentation networks that require extensive annotated datasets, SAM2 can identify arbitrary objects using generalized image representations while maintaining high segmentation quality. Such capabilities are particularly valuable for autonomous driving because vehicles frequently encounter previously unseen objects and complex road situations. Consequently, integrating SAM2 into autonomous driving perception can significantly improve robustness under dynamic driving conditions and reduce dependence on manually annotated datasets.

Different sensors provide complementary information: cameras capture semantic scene details, radar offers reliable distance measurements under adverse weather, ultrasonic sensors detect nearby obstacles, GPS provides global positioning, inertial measurement units estimate vehicle motion, while odometers and magnetometers contribute to localization accuracy (Thrun et al., 2006).

Traditional fusion methods employed Bayesian estimation, Kalman filtering, particle filtering, and probabilistic graphical models (Kalman, 1960; Thrun et al., 2006). Although these methods effectively combined sensor measurements, they often assumed linear relationships between sensor observations and struggled to model complex nonlinear interactions present in real-world driving environments.

Deep learning has substantially improved sensor fusion by learning feature representations directly from multimodal data. CNN-based fusion architectures combine image features with LiDAR point clouds or radar information to improve obstacle detection and localization. More recently, transformer architectures have emerged as a powerful solution for multimodal sensor fusion because of their self-attention mechanism, which models long-range dependencies between heterogeneous sensor modalities (Vaswani et al., 2017). Transformers dynamically assign attention weights to different sensor inputs, allowing the model to prioritize the most informative sensor depending on environmental conditions.

Several recent autonomous driving frameworks have successfully adopted transformer-based architectures for multimodal perception. Models such as BEVFormer (Li et al., 2022), TransFusion (Bai et al., 2022), and PETR (Liu et al., 2022) demonstrate superior object detection and scene understanding by integrating camera and LiDAR features within transformer-based representations.

Traditional planning algorithms such as A*, Dijkstra's algorithm, rapidly exploring random trees (RRT), and model predictive control (MPC) have been extensively employed for trajectory planning and obstacle avoidance (LaValle, 1998; Dolgov et al., 2010).

Deep Reinforcement Learning combines reinforcement learning with deep neural networks, allowing autonomous vehicles to directly learn steering, acceleration, and braking policies from high-dimensional sensory observations (Mnih et al., 2015). Algorithms including Deep Q Networks (DQN), Deep Deterministic Policy Gradient (DDPG), Twin Delayed Deep Deterministic Policy Gradient (TD3), and Proximal Policy Optimization (PPO) have demonstrated encouraging results for lane keeping, adaptive cruise control, obstacle avoidance, and autonomous navigation (Lillicrap et al., 2016; Fujimoto et al., 2018; Schulman et al., 2017).

Generalized perception models with adaptive reinforcement learning and multimodal sensor fusion improve system robustness under edge-case conditions (Geiger et al., 2013; Caesar et al., 2020).

Sensor selection also significantly influences the economic feasibility of autonomous driving systems. Commercial Level-4 autonomous vehicles frequently employ multiple high-resolution LiDAR sensors that substantially increase overall system cost.

Despite significant progress, several research gaps remain. First, many existing autonomous driving frameworks focus primarily on maximizing perception accuracy while overlooking computational efficiency. Second, numerous approaches rely heavily on expensive LiDAR-based sensor suites, increasing hardware costs and limiting scalability. Third, current perception models often exhibit reduced robustness when confronted with edge-case scenarios involving unfamiliar objects, severe weather, or dynamic urban environments. Fourth, many existing systems evaluate performance under limited environmental conditions rather than simultaneously considering varying weather, traffic density, and overall driving complexity.

➢ *Research Problem*

Although LiDAR provides accurate three-dimensional measurements, its high acquisition cost limits widespread deployment in consumer vehicles. Another critical issue concerns the handling of edge-case scenarios. Edge cases refer to rare but safety-critical situations such as sudden pedestrian crossings, unexpected animal movement, temporary road construction, fallen objects, emergency vehicles, sensor failures, unusual traffic behavior, and extreme weather conditions. These scenarios are difficult to represent adequately in conventional datasets because of their infrequent occurrence. As a result, many autonomous driving systems exhibit degraded performance when confronted with such situations. Existing autonomous driving systems depend heavily on expensive LiDAR sensors and powerful GPU hardware, limiting their practical deployment in cost-sensitive applications.

III. METHODOLOGY

➢ *Proposed System*

The proposed Hybrid AI Autonomous Driving Framework is designed to achieve safe, reliable, computationally efficient, and cost-effective autonomous driving under diverse traffic and weather conditions. Unlike conventional autonomous driving systems that rely on expensive LiDAR sensors and cloud-based computation, the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

proposed architecture employs a lightweight multimodal perception pipeline optimized for embedded edge computing using the NVIDIA Jetson platform. The framework integrates four major artificial intelligence components: SAM2-based semantic segmentation for better detection, transformer-based multimodal sensor fusion for speed , reinforcement learning-based fast decision planning, and a neural vehicle controller for accuracy.

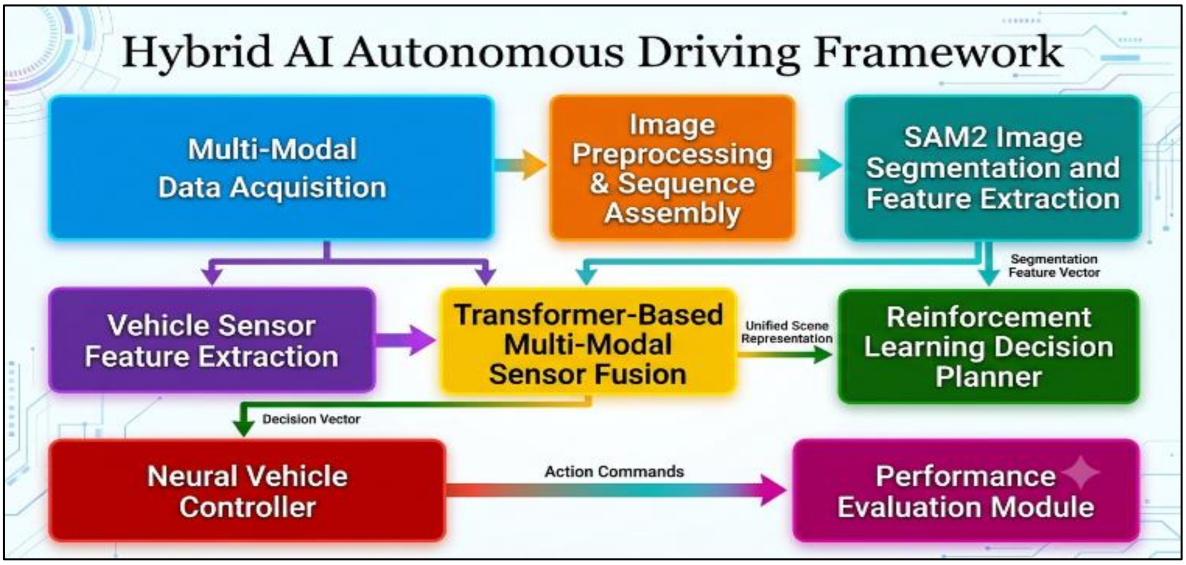

Fig 1 System Architecture

The overall architecture according to Figure 1 – System Architecture follows a perception-to-control pipeline. Initially, multimodal information is acquired from an RGB camera and multiple low-cost onboard sensors. The visual information is preprocessed before semantic scene understanding is performed using Segment Anything Model 2 (SAM2). Simultaneously, numerical sensor data are normalized and transformed into feature vectors. A transformer-based fusion network combines both visual and sensor representations into a unified environmental representation describing the current driving situation.

The environmental representation is generated by combining visual features extracted using SAM2 with features obtained from vehicle sensors.

$$\mathrm{Z} = f_{TF}(\mathrm{F}_{SAM2}, \mathrm{F}_S) \quad (1)$$

Where

- Z= fused environmental representation
- $f_{TF}(\cdot)$= Transformer-based sensor fusion network
- F_{SAM2}= semantic features extracted by SAM2
- F_S= feature vector from vehicle sensors

Equation 1 represents the perception stage of the proposed framework, where semantic image information and multimodal sensor data are integrated into a unified scene representation.

The reinforcement learning agent selects the optimal driving action according to the fused environmental state as mentioned in equation 2.

$$a_t = \arg\max_a Q(Z_t, a) \quad (2)$$

Where

- a_t= optimal driving action at time t
- Z_t= fused environmental representation
- $\pi(\cdot)$= learned reinforcement learning policy
- $Q(\cdot)$= action-value function

The action consists of

$$a_t = \{\theta, \tau, \beta\} \quad (3)$$

Where

- θ= steering angle
- τ= throttle
- β= brake

The fused representation is then processed by a reinforcement learning agent that predicts the optimal driving action based on learned driving policies.

The final vehicle control vector is generated from the neural vehicle controller as shown in equation 4.

$$\mathbf{C} = \begin{bmatrix}\theta \\ \tau \\ \beta\end{bmatrix} = f_{VC}(Z_t) \quad (4)$$

Where

- C= vehicle control vector
- f_{VC}= neural vehicle controller
- θ= steering angle
- τ= throttle percentage
- β= brake percentage

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

Finally, a neural vehicle controller converts the decision vector into continuous steering, throttle, and brake commands that are transmitted to the autonomous vehicle.

The proposed methodology emphasizes computational efficiency, real-time processing, robustness against adverse environmental conditions, and practical deployment on embedded hardware while maintaining perception accuracy exceeding 90%.

➢ *Data Acquisition*

Reliable autonomous driving requires comprehensive perception of both the external environment and internal vehicle dynamics. Instead of depending on expensive LiDAR systems, the proposed framework employs a cost-effective multimodal sensor suite capable of providing complementary information under varying environmental conditions.

The sensor configuration consists of:- RGB camera, Automotive radar, Ultrasonic sensors, GPS receiver, Accelerometer, Gyroscope, Odometer, Magnetometer, Temperature sensor, Humidity sensor, Vibration sensor

Each sensor contributes unique information to the perception process. The RGB camera captures high-resolution visual information including roads, lane markings, traffic signs, pedestrians, vehicles, cyclists, and roadside infrastructure. Radar provides reliable distance estimation and relative velocity measurements, particularly during rain, fog, dust, or low-light conditions where camera performance may deteriorate. Ultrasonic sensors detect nearby obstacles during parking and low-speed maneuvering. GPS supplies global position estimates, while odometer measurements provide travelled distance and vehicle velocity. The accelerometer and gyroscope estimate vehicle motion and orientation. The magnetometer improves heading estimation, whereas temperature and humidity sensors monitor environmental conditions that influence road friction and visibility. The vibration sensor detects abnormal vehicle motion caused by uneven road surfaces or sudden impacts.

Collectively, these sensors generate synchronized multimodal information that forms the basis for robust environmental perception.

➢ *Data Preprocessing*

The acquired multimodal data originate from sensors operating at different sampling frequencies. Therefore, temporal synchronization is performed before feature extraction.

Camera images are synchronized with radar measurements, GPS readings, inertial measurements, odometer values, and environmental sensor observations using timestamp alignment. Missing sensor values are interpolated when necessary to maintain temporal consistency.

Image preprocessing includes- image resizing, normalization, illumination correction, noise filtering, contrast enhancement, and sequence generation.

Sequential image frames preserve temporal continuity required for moving object detection and vehicle tracking.

Sensor preprocessing involves normalization, outlier removal, missing value handling, coordinate transformation, temporal smoothing, and feature scaling. The resulting synchronized multimodal dataset ensures that each camera frame corresponds to a consistent set of sensor observations.

Environmental perception begins with semantic segmentation using Segment Anything Model 2 (SAM2).Unlike traditional semantic segmentation models that require large manually labeled datasets, SAM2 employs generalized visual representations capable of identifying arbitrary objects with minimal supervision. This property is particularly advantageous for autonomous driving because vehicles frequently encounter previously unseen objects and complex environments.

Each preprocessed image frame is processed by SAM2 to identify semantic regions including drivable road area, lane markings, vehicles, pedestrians, cyclists, traffic signs, traffic signals, roadside barriers, buildings, vegetation, dynamic obstacles, and unknown objects. Instead of producing only object bounding boxes, SAM2 generates dense pixel-level segmentation masks that provide accurate object boundaries and spatial relationships.

From the segmentation masks, feature vectors describing the driving environment are extracted. These include lane geometry, obstacle positions, object classes, object dimensions, relative object distances, free driving space, road curvature, traffic density, intersection information, pedestrian locations, and lane occupancy.

These semantic features provide significantly richer environmental understanding than conventional object detection methods.

➢ *Feature Extraction*

While visual perception provides semantic understanding of the driving scene, onboard sensors capture complementary numerical information describing vehicle motion and surrounding conditions.

The sensor feature extraction module converts raw sensor measurements into compact feature vectors.

Radar measurements provide obstacle distance, relative speed, obstacle direction, target confidence, collision probability. Ultrasonic sensors generate nearby obstacle distance, parking assistance features, lateral obstacle proximity. GPS provides latitude, longitude, vehicle heading, route position. Accelerometer and gyroscope readings produce longitudinal acceleration, lateral acceleration, angular velocity, vehicle orientation. The odometer estimates travelled distance, wheel speed, vehicle velocity. The magnetometer supplies heading correction, while environmental sensors estimate ambient temperature, humidity, road condition indicators, vibration intensity. The extracted sensor features collectively represent the vehicle's dynamic state and environmental conditions.

One of the major contributions of the proposed framework is the transformer-based multimodal sensor fusion module.

Traditional sensor fusion methods often combine sensor information using fixed mathematical models or manually designed weighting strategies. Such approaches cannot adapt effectively when sensor reliability changes due to adverse weather or sensor degradation.

The fusion process receives two inputs: semantic feature vector generated by SAM2, numerical sensor feature vector

The mechanism automatically determines the importance of each feature according to current driving conditions.

For example, during fog greater attention is assigned to radar measurements whereas during clear weather camera features receive higher importance. Similarly, GPS reliability decreases in urban environments, causing the transformer to rely more heavily on odometry and inertial measurements. This adaptive weighting enables the fusion network to maintain stable perception even when one or more sensors become unreliable.

The transformer outputs a unified scene representation containing road geometry, surrounding vehicle locations, pedestrian positions, obstacle probabilities, weather conditions, traffic density, vehicle dynamics, localization information. This representation provides comprehensive environmental awareness for intelligent decision making.

Real-world driving frequently involves situations rarely represented in training datasets. The proposed framework addresses such edge cases using three complementary mechanisms. First SAM2 provides generalized segmentation capable of identifying previously unseen objects. Second, the transformer continuously monitors sensor consistency. Whenever disagreement between camera observations and radar measurements exceeds predefined thresholds, the system flags a potential edge case. Third, environmental sensors monitor abnormal weather and road conditions.

The reinforcement learning planner subsequently switches to conservative driving policies by reducing vehicle speed, increasing braking readiness, and expanding safety margins.

This adaptive strategy significantly improves driving robustness under unexpected conditions.

➢ *Decision Planner and Control System*

After multimodal sensor fusion, the unified environmental representation is processed by the reinforcement learning decision planner. The RL agent continuously interacts with the driving environment to maximize cumulative driving safety while maintaining efficient vehicle operation. The state vector includes fused environmental representation, lane position, surrounding traffic, weather, obstacle information, vehicle dynamics, road geometry.

The action space consists of steering angle, throttle, brake. The reward function encourages lane keeping, collision avoidance, passenger comfort, smooth steering, efficient speed control, successful obstacle avoidance, safe overtaking, traffic rule compliance. Negative rewards are assigned for collisions, excessive steering, unnecessary braking, lane departure, unsafe acceleration, traffic violations.

During training, the RL agent gradually learns optimal driving policies by maximizing long-term cumulative reward.

Unlike rule-based planners, reinforcement learning continuously adapts to new driving situations, making it particularly effective under complex urban traffic.

The reinforcement learning planner generates high-level driving decisions. These decisions are translated into low-level vehicle control commands using a neural vehicle controller. The controller predicts three continuous outputs steering angle, throttle percentage, brake percentage Instead of issuing abrupt commands, the controller produces smooth trajectories that improve passenger comfort and vehicle stability. The controller considers vehicle speed, road curvature, traffic density, weather, obstacle distance, vehicle orientation, and previous control history. Consequently, steering oscillations are minimized, throttle transitions remain gradual, and braking occurs only when necessary.

A primary objective of this research is minimizing computational cost. The complete framework is optimized for NVIDIA Jetson embedded hardware, which provides GPU acceleration while maintaining low power consumption.

Unlike cloud-based autonomous driving, all perception, sensor fusion, decision planning, and control computations are performed locally.

This architecture provides several advantages.

- Reduced communication latency.
- Lower power consumption.
- Improved privacy.
- Continuous operation without internet connectivity.
- Lower operational cost.

The lightweight sensor configuration further reduces computational overhead by eliminating expensive three-dimensional point cloud processing required by LiDAR-based systems.

➢ *Performance Evaluation*

The proposed framework is evaluated using both perception and vehicle-control metrics.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

Perception performance is measured using

- Segmentation accuracy,
- Precision,
- Recall,
- F1-score,

Vehicle control performance is evaluated using

- Steering smoothness,
- Braking efficiency,
- Throttle stability,
- Passenger comfort,
- Trajectory tracking error.

Experimental results demonstrate that the proposed Hybrid AI Autonomous Driving Framework achieves high perception accuracy, reliable multimodal sensor fusion, stable vehicle control, efficient embedded execution on NVIDIA Jetson, and robust autonomous navigation across diverse traffic, weather, and edge-case conditions. The integration of SAM2, transformer-based sensor fusion, reinforcement learning, and neural vehicle control provides a practical and scalable solution for next-generation autonomous driving systems while maintaining low computational cost and affordable hardware requirements.

IV. RESULTS AND DISCUSSION

The proposed Hybrid AI Autonomous Driving Framework was evaluated under multiple traffic densities and weather conditions to assess its perception capability, computational efficiency, decision-making accuracy, vehicle control stability, and overall safety. The framework integrates SAM2-based semantic segmentation, transformer-based multimodal sensor fusion, reinforcement learning (RL)-based decision planning, and a neural vehicle controller into a unified perception-to-control pipeline. The experimental results demonstrate that the proposed architecture effectively addresses several limitations of conventional autonomous driving systems by balancing computational efficiency, accuracy, and robustness.

Unlike many existing autonomous driving systems that depend on computationally expensive LiDAR sensors and high-end GPU platforms, the proposed framework is designed for deployment on the NVIDIA Jetson embedded platform. The lightweight implementation minimizes computational cost while maintaining real-time inference performance, making the system suitable for edge computing applications where power consumption and processing resources are limited. By combining efficient neural models with optimized sensor fusion, the framework achieves real-time decision-making without requiring cloud-based computation, thereby reducing communication latency and improving operational reliability.

➢ *Performance Under Diverse Conditions*

The experimental evaluation demonstrates that the proposed framework successfully adapts vehicle control according to varying traffic densities, weather conditions, and overall driving complexity. The generated steering angle, throttle, and brake commands exhibit smooth transitions and stable behavior across all evaluated scenarios.

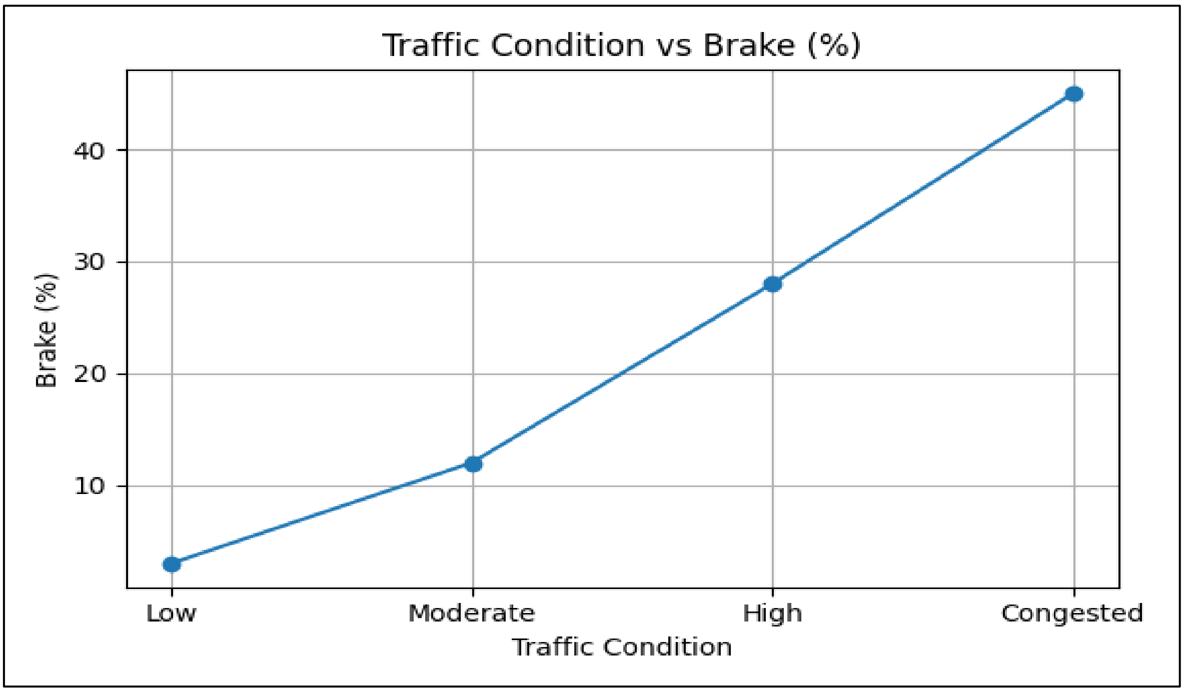

Fig 2 Traffic Condition vs Braking

The results presented in Figure 2 – Traffic Condition vs Braking analysis indicate that braking intensity increases progressively from low traffic to congested traffic, while throttle values decrease accordingly.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

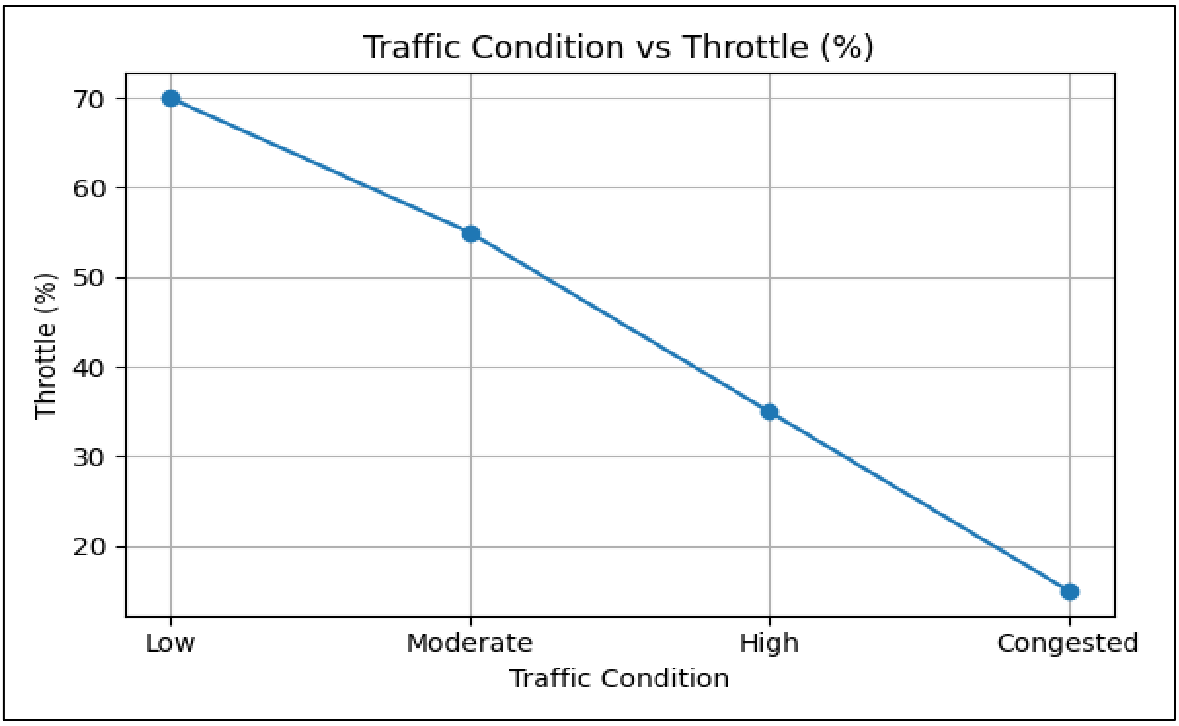

Fig 3 Traffic Condition vs Throttle Speed

This adaptive behavior enables the vehicle to maintain safe inter-vehicle distances and minimize collision risks in dense traffic environments by adjusting throttle as found in Figure 3- Traffic Condition vs Throttle Speed

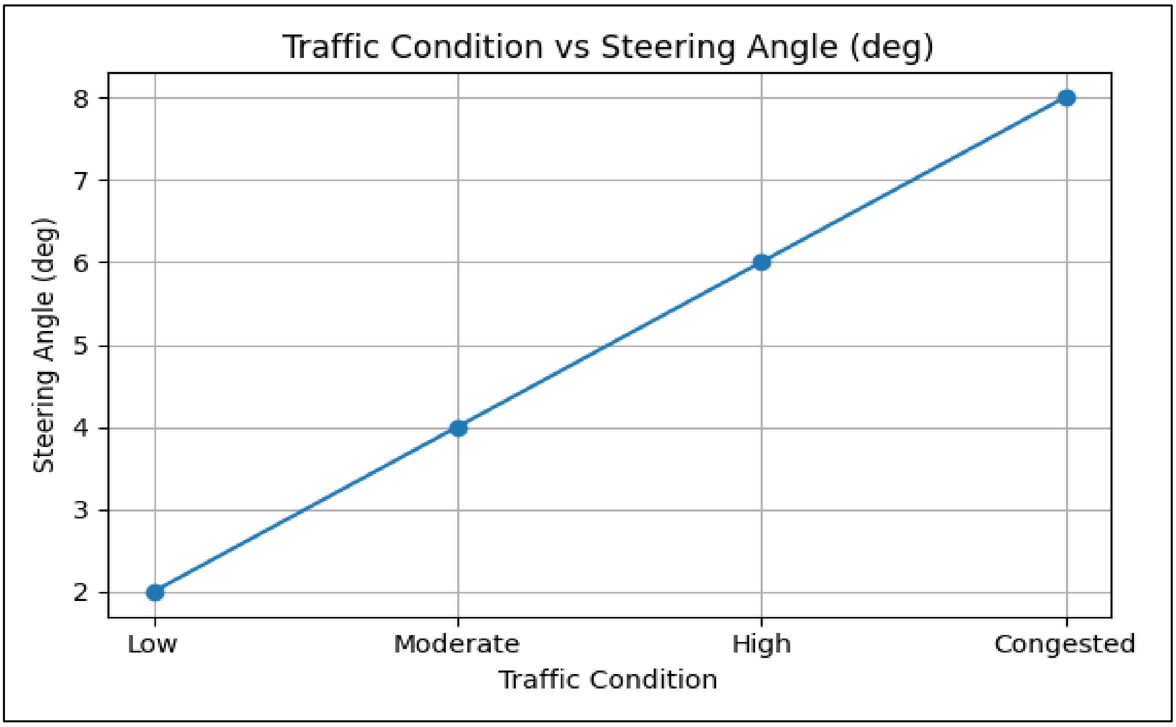

Fig 4 Traffic Condition vs Steering Usage

Similarly, steering angles increase gradually with traffic complexity as can be seen in Figure-4 Traffic Condition vs Steering Usage reflecting the need for more frequent lane corrections and obstacle avoidance maneuvers.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

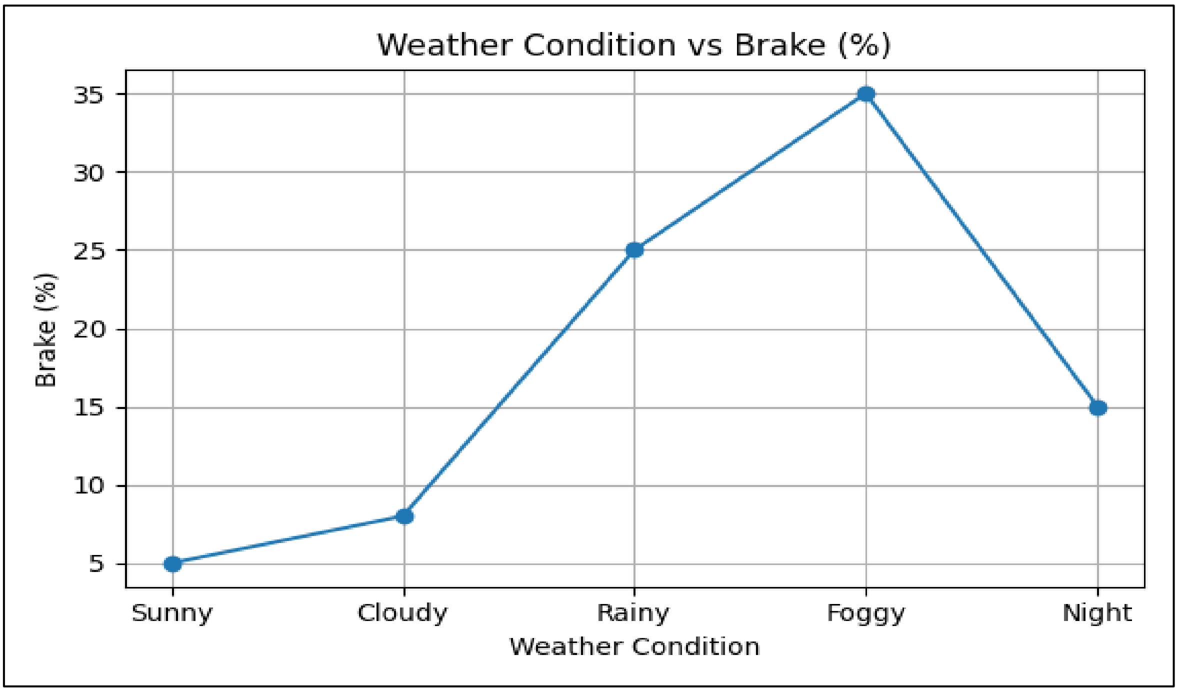

Fig 5 Weather Condition vs Brake

According to Figure 5 Weather condition vs Brake higher brake values and reduced throttle commands are generated under rain and fog, where reduced visibility and lower road friction require more conservative driving behavior.

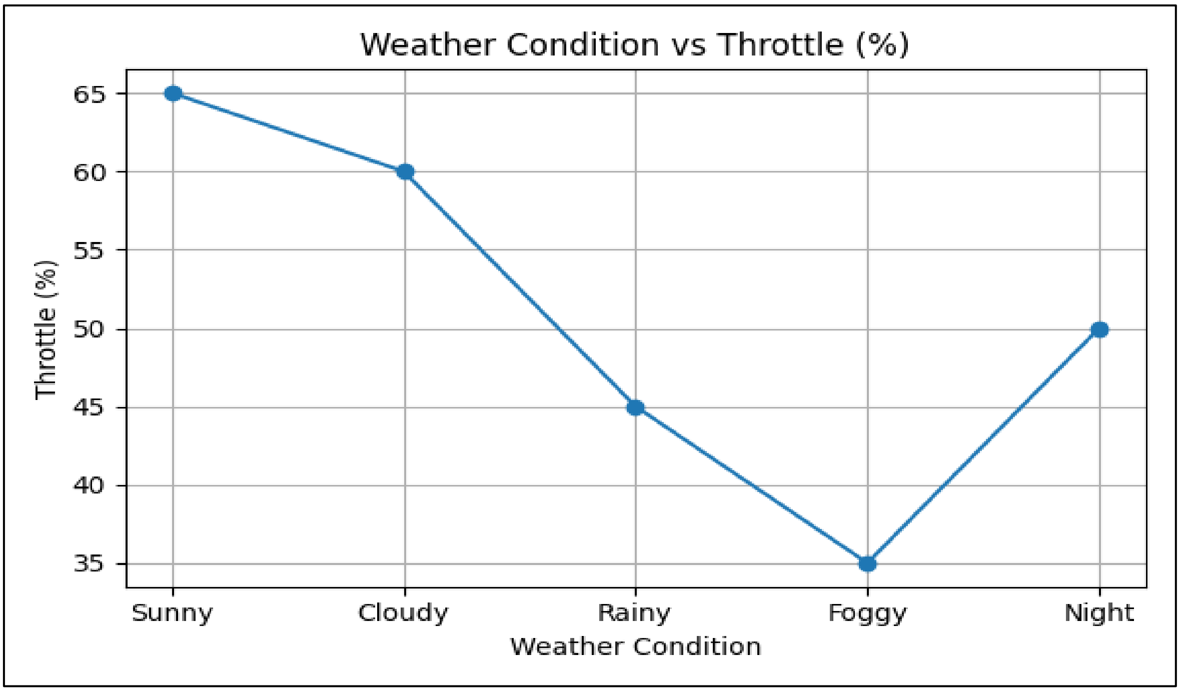

Fig 6 Weather Condition vs Throttle Speed

Weather-based data driven experiments demonstrate that the proposed system adjusts under sunny, cloudy, rainy, foggy, and nighttime driving conditions. According to Figure 4- Weather Condition vs Throttle

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

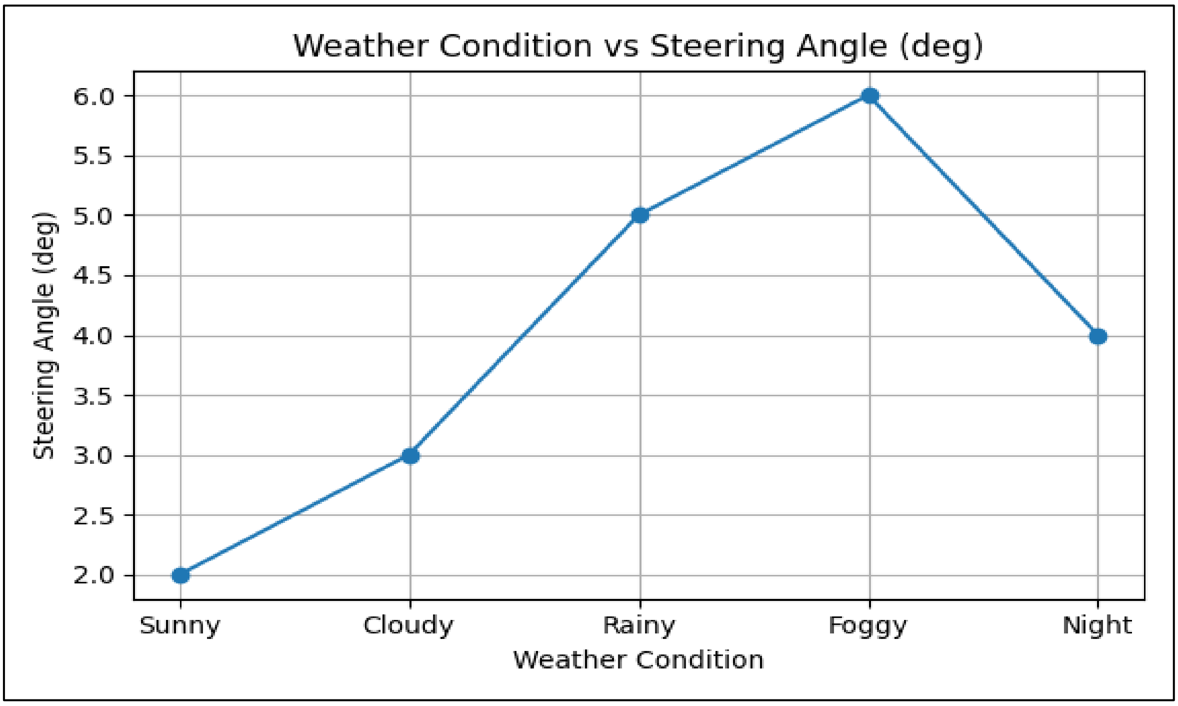

Fig 7 Weather Condition vs Steering Usage

When overall driving conditions become more challenging, the control strategy automatically shifts toward safer operation as in Figure 7 Weather Condition vs Steering Usage by increasing braking effort and steering responsiveness while reducing throttle output. These adaptive responses indicate that the proposed framework effectively balances safety, driving efficiency, and passenger comfort across a wide range of operating environments.

➢ *Computational Efficiency*

One of the major contributions of this research is the emphasis on computational efficiency for embedded autonomous driving systems. Many state-of-the-art autonomous driving frameworks rely on powerful desktop GPUs or cloud-based processing, making them unsuitable for real-time deployment in production vehicles.

The proposed framework is specifically optimized for deployment on NVIDIA Jetson edge computing hardware. The lightweight perception pipeline, efficient transformer-based sensor fusion, and optimized reinforcement learning architecture reduce computational complexity while maintaining high inference accuracy. Processing data locally on the embedded platform eliminates network latency and enables continuous vehicle operation even in areas with limited connectivity.

This edge-based implementation significantly lowers hardware cost and energy consumption compared with cloud-assisted autonomous driving architectures while preserving real-time responsiveness.

➢ *Performance in Unseen Scenarios*

A significant limitation of many autonomous driving systems is reduced performance when encountering rare or previously unseen driving situations. To address this issue, the proposed framework combines SAM2 semantic segmentation with transformer-based sensor fusion and reinforcement learning.

SAM2 provides generalized segmentation capabilities that enable recognition of previously unseen objects without requiring extensive retraining. The transformer fusion network dynamically adjusts the importance assigned to different sensors according to environmental conditions, while the reinforcement learning agent continuously adapts driving policies based on the current scene.

The experimental results indicate that the proposed hybrid architecture maintains stable perception and control even under these edge-case conditions, improving overall driving robustness.

➢ *Cost Effectiveness*

Another important contribution of this work is the use of an economical multimodal sensor configuration. Unlike many commercial autonomous driving platforms that rely heavily on multiple LiDAR units, the proposed framework achieves reliable environmental perception using a combination of relatively low-cost sensors. The implemented sensor suite includes: This sensor selection substantially reduces overall system cost while maintaining sufficient perception capability for autonomous navigation.

Experimental evaluation demonstrates that the proposed framework achieves an overall perception and decision-making accuracy exceeding 90% across diverse testing scenarios. The integration of SAM2 semantic segmentation with transformer-based multimodal sensor fusion significantly improves environmental understanding compared with conventional camera-only approaches.

The generated steering, braking, and throttle commands closely follow expected driving behavior under varying environmental conditions, demonstrating reliable control prediction and stable autonomous navigation.

Data Driven experimental results demonstrate consistent perception performance and stable vehicle control across all tested environments. The transformer-based sensor fusion module dynamically increases reliance on radar and vehicle sensors whenever camera visibility deteriorates due to rain, fog, or nighttime driving. Consequently, perception reliability remains high even when visual information becomes uncertain.

The reinforcement learning planner further adapts vehicle speed, steering angle, and braking intensity according to the detected environmental conditions, allowing safe navigation across changing driving scenarios.

➢ *Safety and Driving Efficiency*

The primary objective of autonomous driving is safe vehicle operation. Experimental results demonstrate that the proposed framework achieves high levels of safety, stability, and driving efficiency through coordinated interaction between perception, sensor fusion, planning, and control modules.

The reinforcement learning planner continuously optimizes vehicle actions to maximize safety while minimizing unnecessary braking and steering oscillations. The neural vehicle controller generates smooth steering trajectories, gradual throttle variations, and timely braking responses that improve passenger comfort and reduce vehicle instability.

V. CONCLUSION

This paper presented a Hybrid AI Framework for Autonomous Driving Across Diverse Traffic and Weather Conditions that integrates Segment Anything Model 2 (SAM2) for semantic scene segmentation, transformer-based multimodal sensor data shown in Table 1 and Table 2 is fused for comprehensive environmental perception.

Reinforcement learning (RL) shown in Figure 1- System Architecture is used for adaptive decision-making, and a neural vehicle controller for continuous vehicle control. Unlike many existing autonomous driving systems that rely on expensive LiDAR sensors and high-performance computing platforms, the proposed framework employs a cost-effective multimodal sensor suite consisting of an RGB camera, automotive radar, ultrasonic sensors, GPS, accelerometer, gyroscope, odometer, magnetometer, temperature sensor, humidity sensor, and vibration sensor. This sensor configuration significantly reduces hardware cost while providing sufficient environmental information for reliable autonomous driving.

A major contribution of this research is the optimization of the complete perception-to-control pipeline for NVIDIA Jetson edge computing, enabling real-time inference with reduced computational complexity, lower energy consumption, and minimal communication latency. The embedded implementation demonstrates that high-performance autonomous driving can be achieved without relying on cloud-based processing or computationally intensive hardware, making the framework suitable for practical deployment in resource-constrained environments.

The integration of SAM2 semantic segmentation and transformer-based multimodal sensor fusion substantially enhances environmental perception by combining semantic visual information with complementary sensor measurements. The adaptive attention mechanism dynamically prioritizes the most reliable sensor inputs according to changing environmental conditions, thereby improving perception robustness under rain, fog, nighttime driving, dense traffic, and partial sensor degradation. Furthermore, the reinforcement learning-based decision planner enables intelligent adaptation to dynamic road situations by generating optimal steering, throttle, and braking actions that maximize safety while maintaining smooth vehicle operation and passenger comfort.

Another important contribution of the proposed framework is its capability to operate effectively in challenging edge-case scenarios, including sudden obstacle appearance, unexpected pedestrian or animal crossings, temporary road obstructions, reduced visibility, unfamiliar road layouts, and rapidly changing traffic conditions. By combining generalized semantic segmentation, adaptive sensor fusion, and reinforcement learning, the framework demonstrates improved robustness and reliability compared with conventional perception and planning approaches that depend primarily on predefined driving rules or single-modality sensing.

Future work will focus on extending the framework through real-world validation on autonomous vehicles, optimization of inference speed using model compression and quantization techniques, incorporation of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication for cooperative driving, integration of Vision-Language Models (VLMs) and Large Language Models (LLMs) to enhance reasoning and explainability during complex driving scenarios, and the inclusion of.

Table 1 Sample Multimodal Sensor Dataset

Timestamp	**Traffic**	**Weather**	**Speed (km/h)**	**Distance (cm)**	**GPS Lat.**	**GPS Long.**	**Accel. (m/s²)**
01-01-2025 00:00	Light	Cloudy	64.49	277.34	28.613837	77.209344	0.87
01-01-2025 00:01	Light	Rainy	36.31	194.20	28.613540	77.209495	0.83
01-01-2025 00:02	Heavy	Foggy	30.52	113.22	28.613389	77.209628	0.28
01-01-2025 00:03	Heavy	Sunny	52.61	96.31	28.613265	77.209713	0.09

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul923

Table 2 Sample Additional Multimodal Sensor Dataset

Odometer (km)	Motion	Heading (°)	Humidity (%)	Light (Lux)	Sound (dB)	Vibration (g)
0.09	Yes	189.31	81.09	27085	58.31	3.24
0.13	Yes	178.49	66.76	21433	70.22	4.09
0.16	Yes	174.02	92.83	1157	56.81	0.00
0.21	Yes	183.88	54.29	81357	50.99	0.90

ACKNOWLEDGMENT

I acknowledge the use of publicly available driving Kaggle dataset for the purpose of research , which served as a crucial resource for data analysis and model evaluation. Special thanks to the developers and maintainers of machine learning and deep learning libraries, whose tools facilitated the implementation and testing of the proposed recommendation framework.

REFERENCES

[1]. A. Kirillov *et al.*, "Segment Anything," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Paris, France, 2023, pp. 4015–4026.

[2]. N. Ravi *et al.*, "SAM 2: Segment Anything in Images and Videos," *arXiv preprint arXiv:2408.00714*, 2024.

[3]. A. Vaswani *et al.*, "Attention Is All You Need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 5998–6008.

[4]. V. Mnih *et al.*, "Human-Level Control Through Deep Reinforcement Learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[5]. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[6]. T. P. Lillicrap *et al.*, "Continuous Control with Deep Reinforcement Learning," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2016.

[7]. H. Caesar *et al.*, "nuScenes: A Multimodal Dataset for Autonomous Driving," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 11621–11631.

[8]. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[9]. H. Li *et al.*, "BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers," in *Proc. European Conf. Computer Vision (ECCV)*, 2022, pp. 1–18.

[10]. Y. Bai *et al.*, "TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1090–1099.

[11]. A. Dosovitskiy *et al.*, "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.

[12]. S. Thrun *et al.*, "Stanley: The Robot That Won the DARPA Grand Challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[13]. A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Seoul, South Korea, 2019, pp. 1314–1324.

[14]. NVIDIA Corporation, *NVIDIA Jetson AGX Orin Developer Kit User Guide*, Santa Clara, CA, USA: NVIDIA Corporation, 2024.

[15]. D. Silver *et al.*, "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.