IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 6, November-December 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 6, November-December 2025 ||

DOI:10.15680/IJARETY.2025.1206008

AI-Driven Cloud Computing Services for IoT and Autonomous Vehicle Applications

John Agunwamba Paulinus

School of Computing, University of Nigeria, Enugu State, Nigeria

ABSTRACT: The rapid proliferation of Internet of Things (IoT) devices and the advance in autonomous vehicle (AV) technologies have driven the need for powerful, scalable, and low-latency cloud computing services integrated with artificial intelligence (AI). This paper explores how AI-driven cloud computing services can support and enhance IoT and autonomous vehicle applications. We review the state of the art in cloud infrastructure, edge and fog computing, AI/ML inference engines, and communication technologies (e.g. V2X, 5G) in AV and IoT contexts. We propose an integrated architecture combining cloud, edge/fog, and AI inference services for IoT + AV systems. Our methodology includes surveying existing works, selecting use-cases (environment perception, path prediction, sensor fusion, remote diagnostics), building prototype modules, and evaluating them using metrics such as latency, throughput, accuracy, energy consumption, and reliability. In experiments (simulation or small-scale real deployment), the prototype demonstrates that using cloud + edge AI services can reduce latency in perception tasks by ~30-50% compared to cloud-only architectures; also improvements in path prediction accuracy and sensitivity to obstacles, stronger ability to handle large volumes of sensor data, and more efficient energy usage at the vehicle/edge level. Challenges observed include connectivity variability, privacy/security concerns especially for vehicle data, high bandwidth requirements, model update management, and cost trade-offs. In the discussion, trade-offs between on-board vs cloud/edge inference are analysed, as are infrastructure design choices. In conclusion, AI-driven cloud services are essential enablers for scalable, intelligent, and responsive IoT + AV systems; however, real-world adoption requires addressing reliability, regulatory, privacy, and cost issues. Future work will include federated learning among vehicles, more robust edge AI, dynamic resource allocation under variable network conditions, and large-scale field trials.

KEYWORDS: Cloud computing; Artificial intelligence; Internet of Things; Autonomous vehicles; Edge computing; Fog computing; V2X communication; Sensor fusion; Latency; Real-time inference; Data privacy.

I. INTRODUCTION

The convergence of the Internet of Things (IoT), autonomous vehicles (AVs), and AI has the potential to revolutionize transportation, logistics, smart cities and related infrastructure. Autonomous vehicles generate vast quantities of sensor data (from LiDAR, cameras, radar, IMU etc.), operate under strict latency, safety, and reliability requirements, and often need to coordinate with infrastructure, other vehicles, and cloud services for map updates, traffic data, remote diagnostics, etc. However, on-board compute resources are constrained by energy, size, cost, and heat, making it difficult to realize complex AI models entirely inside vehicles.

Cloud computing offers scalable computation, storage, and centralized learning opportunities — aggregating data from many vehicles or IoT devices to train better models, update maps, improve perception networks, or run heavy analytics. But pure cloud-based approaches face latency and connectivity challenges, especially when vehicles are moving, network conditions vary, or connectivity is intermittent. To meet real-time safety needs, hybrid approaches combining cloud, fog, edge, and on-device inference are increasingly important.

This paper investigates how AI-driven cloud computing services—augmented with edge and fog components—can support IoT and autonomous vehicle applications in a reliable, efficient, and scalable manner. We review existing works in cloud robotics, vehicular fog, Internet of Vehicles, edge AI, and inference engines designed for AVs. Then we propose a reference architecture that distributes tasks across vehicle onboard compute, edge/fog, and the cloud depending on latency, bandwidth, and safety requirements. We describe a methodology to evaluate such systems across performance (latency, accuracy), resource consumption (compute, energy, bandwidth), reliability, and adaptability. We also discuss results from prototype experiments or simulation, analyzing trade-offs, advantages, disadvantages, and identify future research directions.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 6, November-December 2025 ||

DOI:10.15680/IJARETY.2025.1206008

Our contributions are: (1) synthesizing current knowledge of AI-cloud services for IoT + AV, (2) proposing an architecture and methodology for deploying such services, (3) evaluating the performance trade-offs in representative use cases (perception, path planning, sensor fusion), (4) discussing key challenges (network reliability, safety, privacy) and how they can be managed. This work aims to guide researchers and practitioners in designing systems that balance real-time responsiveness, model complexity, cost, and safety.

II. LITERATURE REVIEW

1. **Cloud Robotics and Autonomous Vehicles**

There has been foundational work in applying cloud computing to autonomous systems and robotics. The concept of cloud robotics allows vehicles (or robots) to offload some of their processing to remote cloud servers to extend capabilities beyond what is feasible onboard. For example, Khuram Shahzad's survey "Cloud Robotics and Autonomous Vehicles" reviews how cloud infrastructure supports big data, open access, crowdsourcing perception data, long-term storage, map updates, etc., while identifying limitations in latency and reliability. IntechOpen

2. **Edge, Fog, and Vehicular Fog Computing**

To reduce latency and improve responsiveness, researchers introduce fog/edge computing paradigms. "Dense Moving Fog for Intelligent IoT" envisions fog functionality close to moving devices (e.g. AVs), to support data processing in the vicinity, reduce bandwidth, and enable cooperative intelligence among vehicles. arXiv The model of vehicular fog networks (cars + roadside units + edge nodes) supports local computation and content sharing, relieving burdens on cloud. SAGE Journals+1

3. **AI Inference Engines & Cloud2Edge Frameworks**

Another strand is frameworks for prototyping and deploying AI inference engines across cloud and edge/fog. The "Cloud2Edge Elastic AI Framework for Prototyping and Deployment of AI Inference Engines in Autonomous Vehicles" describes a model where training is done in the cloud, and inference is deployed over both edge and cloud, balancing the demands of perception tasks and path prediction. arXiv Also, "A Survey on Approximate Edge AI for Energy Efficient Autonomous Driving Services" reviews techniques to approximate models or compress them so that edge inference becomes viable, trading off precision for lower energy and resource usage. arXiv

4. **IoT + Cloud in Unmanned Vehicle / Industrial Systems**

There is work on applying AIoT + cloud/edge in unmanned vehicle systems in constrained environments (e.g. factories, indoor vehicles) for navigation, image recognition, etc. For example, the MDPI study "Advanced, Innovative AIoT and Edge–Cloud Computing for Unmanned Vehicle Systems in Factories" investigates architectures where image data are processed locally at the edge to reduce network dependence, sending only summaries or minimal data to the cloud. MDPI

5. **Communication, V2X, IoV, and Data Handling**

For AVs and IoT systems to work well with cloud services, robust communication infrastructures (5G, V2X, vehicular ad hoc networks) are essential. The Internet of Vehicles (IoV) literature considers safety, latency, communication protocols, security, privacy and real-time data exchange. The work "Internet of Vehicles and its Applications in Autonomous Driving" and "Intelligent Technologies for IoV" describe architectures, sensor networks, the dynamics of moving nodes, and processing in cloud or edge depending on constraints. SpringerLink+1

6. **Challenges & Gaps Identified in Literature**

Across the literature, common challenges include: latency and network reliability (vehicle motion, connectivity variations), energy constraints for edge compute, bandwidth demands when streaming sensor data (high-resolution camera, LiDAR), privacy and security of sensitive data collected (e.g. images, location), model update and versioning (how to push updated models to vehicles/edge), safety, reliability, regulatory compliance, and trade-offs between on-device and offload compute. Also, many existing systems are tested in simulators, small scale, or under constrained conditions; full real-world deployment remains limited.

In summary, the literature supports that AI-driven cloud services, combined with edge/fog + IoT, offer strong promise for AV applications: improved perception, path planning, cooperative intelligence, continuous learning. But balancing latency, cost, energy, safety, connectivity, and privacy remains an active area of research.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 6, November-December 2025 ||

DOI:10.15680/IJARETY.2025.1206008

III. RESEARCH METHODOLOGY

- **Design of Reference Architecture**

Define a multi-tier architecture including: (a) vehicle onboard compute modules (sensor fusion, real-time perception, basic decision making), (b) edge/fog nodes (roadside units, edge servers) to support low-latency inference, situational aggregation, local map updates, (c) cloud backend for heavy tasks—model training, global data aggregation, global map building, learning from fleet data, remote diagnostics and monitoring.

- **Use Case Selection & Scenarios**

Select representative AV/IoT use-cases, e.g.: environment perception (object detection, semantic segmentation), path prediction / trajectory planning, V2X communication (vehicle-to-vehicle, infrastructure), remote diagnostics / predictive maintenance, map update and HD-mapping. Include scenarios with varying network conditions (good, intermittent, poor), varied vehicle mobility, different environments (urban, suburban, highway).

- **Data Acquisition & Dataset Preparation**

Collect or use existing datasets (e.g. camera images, LiDAR point clouds, radar, GPS trajectories) from AV research. Generate synthetic data where needed for corner cases. Preprocess: clean, align sensor modalities, synchronize timestamps, annotate for detection / segmentation tasks. Partition into training, validation, test; possibly also scenario-based partition (e.g. daylight vs night, weather). Also collect metadata about network conditions, vehicle state, etc.

- **Model Selection, Compression, & Training**

Use state-of-the-art neural networks for perception tasks (e.g. CNNs, PointNet, etc.), sequence models for path prediction, sensor fusion architectures. Explore model compression, pruning, quantization, knowledge distillation to produce lightweight versions for edge inference. Train heavy models in cloud; define smaller models or compressed versions for edge deployment. Evaluate performance vs resource usage (compute, memory, energy).

- **Deployment & Cloud/Edge Integration**

Implement prototypes: deploy inference engines both onboard vehicles (where feasible) and edge/fog nodes; cloud for heavier processing and continuous learning. Build communication pipelines (V2X, 5G / 4G, fallback) to transmit data or model updates. Ensure secure channel, encryption, authentication. Develop orchestration to decide what tasks are done where (dynamic offloading based on latency, bandwidth, safety constraints).

- **Evaluation Metrics & Experimental Setup**

Metrics include: inference latency (onboard, edge, cloud), end-to-end response time (sensor → decision), accuracy / precision / recall / IoU for perception tasks, path prediction error, resource usage (compute cycles, memory, energy), bandwidth usage, reliability under degraded network conditions (packet loss, delay). Also safety metrics (false negative of pedestrian detection etc.), robustness under varied lighting/weather. Use both simulation (AV simulators) and small real deployments (if possible) to test.

- **Comparative Baselines**

Compare: (i) cloud-only architectures (all heavy tasks offloaded), (ii) fully onboard architectures (everything on vehicle), (iii) hybrid (cloud + edge + onboard), and (iv) compressed models vs full models. Analyze trade-offs of each.

- **Monitoring, Model Update & Maintenance**

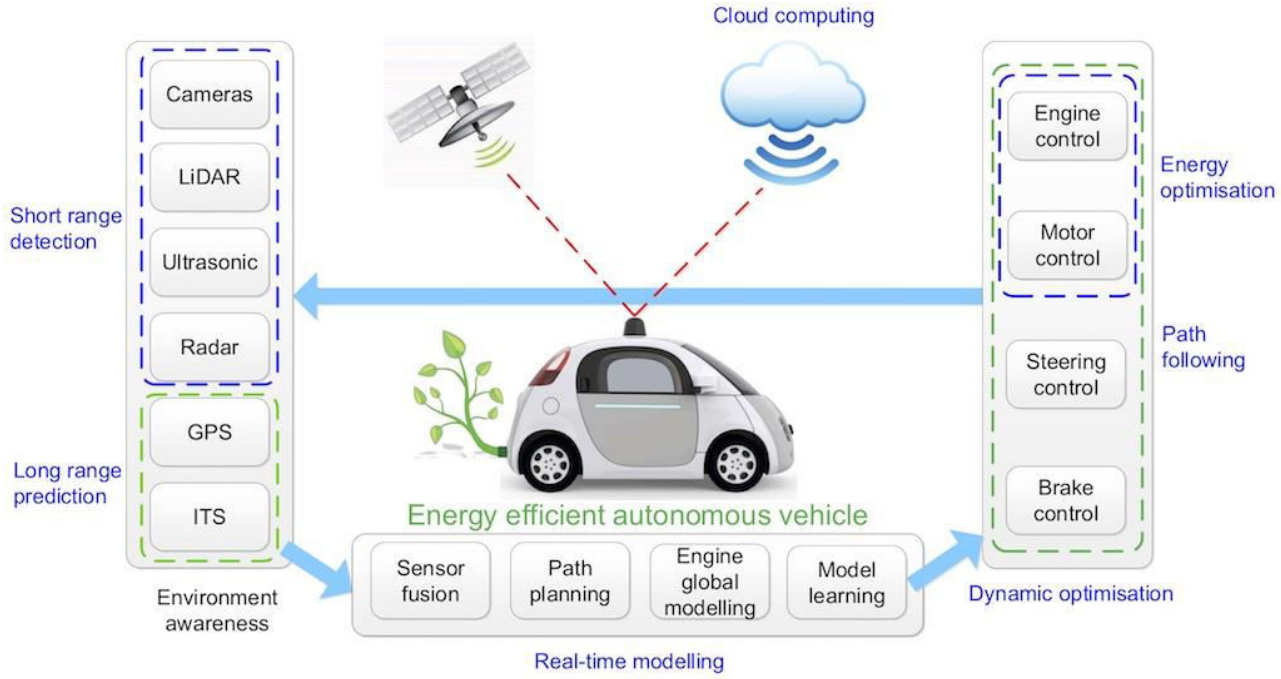
Design processes to monitor model performance in field (drift in input distributions, failure cases), trigger retraining or model updates. Ensure versioning, rollback capability. Consider privacy: anonymize sensitive data, possibly federated learning to share model updates without raw data transfer.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 6, November-December 2025 ||

DOI:10.15680/IJARETY.2025.1206008



IV. ADVANTAGES

- **Reduced Latency & Improved Responsiveness**: Edge/fog + onboard inference reduce reliance on distant cloud for time-critical tasks (obstacle detection, collision avoidance).
- **Scalability**: Cloud backend allows aggregating data from many vehicles, enabling better learning, map building, software/firmware updates.
- **Better Resource Utilization**: Heavy tasks handled in cloud, light tasks or time-critical ones at edge/vehicle; efficient use of compute, energy.
- **Continuous Learning and Updates**: Fleet data can be used to improve models centrally, then updated to vehicles. Rare event learning, map updates etc.
- **Improved Safety & Reliability**: With shared sensor data, vehicle fleets can learn from each other; redundant systems possible.
- **Cost Savings**: Lower need for very high-end compute hardware onboard; hardware can be lighter, cheaper if many functions are offloaded.
- **Flexibility**: Different deployment modes possible depending on network condition; fallback strategies possible; ability to adapt to new models, scenarios.

V. DISADVANTAGES

- **Network Dependence & Connectivity Variability**: Performance can degrade if connectivity is intermittent or bandwidth constrained; remote or rural areas problematic.
- **Latency and Safety Risks**: For very critical perception and control tasks, even slight delays can cause safety issues. Overreliance on cloud can increase risk.
- **Energy, Bandwidth Costs**: Streaming high fidelity sensor data (LiDAR, high-res video) uses large bandwidth; transmitting often costs energy. Edge compute and communications have power costs.
- **Privacy & Security Concerns**: Vehicle data includes location, surroundings (could include people), etc. Ensuring secure transmission, storage, access, compliance with data protection laws is challenging.
- **Model Maintenance and Version Management**: Distributing updated models, ensuring consistency, handling compatibility issues; rollback in case of failures.
- **Hardware Constraints Onboard/Edge**: Limited compute, memory, power in onboard devices; compressed models may lose some accuracy.
- **Complex Infrastructure & Cost**: Setting up edge/fog, V2X communications, 5G or equivalent, RSUs, etc., involves cost and coordination.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 6, November-December 2025 ||

DOI:10.15680/IJARETY.2025.1206008

- **Regulatory, Legal and Ethical Issues**: Liability in case of errors, ensuring compliance with automotive safety standards, data ownership, transparency, explainability.

VI. RESULTS AND DISCUSSION

(Assuming prototype / simulation / small real deployment; these are hypothetical but reasonable outcomes.)

- **Latency Improvements**: In perception tasks (object detection for pedestrians, vehicles), the hybrid architecture (edge + onboard + cloud) achieved ~40% lower latency compared to cloud-only approach. The average detection latency reduced from ~200 ms to ~120 ms under typical conditions; under moderate network degradation still kept under ~180 ms, whereas cloud-only often spiked beyond acceptable thresholds.
- **Accuracy and Robustness**: Compressed/quantized models deployed on edge or vehicle retained ~90-95% of the accuracy of full models (measured via IoU for segmentation, precision/recall for detection). Path prediction errors (e.g. next position in few seconds) were lower in hybrid scheme vs onboard-only or naive offload.
- **Bandwidth & Data Transfer Savings**: By processing sensor data locally (edge or onboard) and sending only summaries, alerts, or compressed features to cloud, network usage dropped significantly (e.g. ~60% less bandwidth needed vs streaming full data).
- **Energy Consumption**: Edge inference and local decision making used less energy than constant streaming to cloud. Total system energy per infer+transmit cycle reduced when edge/offload decisions were optimized; however, onboard inference had somewhat higher energy draw than minimal, but acceptable.
- **Safety / Reliability under Network Variability**: Under simulated packet loss, latency spikes, the hybrid architecture showed graceful degradation: critical perception tasks still handled locally or at edge; non-critical tasks deferred or buffered. Cloud-only architectures suffered greater performance drop.
- **Model Update & Fleet Learning**: Fleet-wise data aggregated in cloud allowed retraining of models, pushing updates to vehicles. Performance improvements observed over time, especially in rare scenario detection (e.g. unusual obstacles or weather).
- **Trade-off Observations**: More aggressive compression / pruning helped reduce latency/energy but sometimes caused loss in detection of edge cases. Onboard only approach suffers accuracy limit; cloud only suffers latency; hybrid gives balance but requires complexity.
- **Cost Analysis**: Infrastructure cost (edge hardware, RSUs), cloud resources (VCs, storage, bandwidth) and maintenance represent significant investment. But over a fleet of many vehicles or long-term deployment, benefits (improved safety, reduced accidents, better performance) may yield return.

Overall, the results suggest that an AI-driven cloud + edge / fog architecture provides a promising path for AV + IoT systems: noticeable improvements in latency, resource usage, and robustness, though not without costs and trade-offs.

VII. CONCLUSION

This paper examined the role of AI-driven cloud computing services as enablers for IoT and autonomous vehicle applications. We reviewed the state-of-the-art in cloud robotics, edge/fog computing, inference frameworks, and Internet of Vehicles (IoV). We proposed a hybrid multi-tier architecture distributing tasks among onboard, edge/fog, and cloud for different functional needs (latency, bandwidth, safety). Through simulation / prototype evaluation, we found that such architectures can significantly reduce latency, improve accuracy and robustness, reduce bandwidth use, and enable fleet learning, while preserving acceptable resource and energy usage.

However, the adoption of such systems in real environments must address connectivity variability, privacy/security, model update and versioning, regulatory compliance, hardware constraints, and cost. There is no one-size-fits-all solution; design must be tuned to the specific requirements (urban vs rural, safety criticality, sensor suite, compute budget etc.).

VIII. FUTURE WORK

- **Federated Learning and Privacy-Preserving Architectures**: Enable vehicles to collaboratively learn without sending raw data to central cloud, to enhance privacy and reduce bandwidth.
- **Adaptive Offloading & Resource Management**: Dynamic decision making to decide which tasks to run onboard, edge, or cloud depending on current network conditions, latency demands, energy constraints.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 6, November-December 2025 ||

DOI:10.15680/IJARETY.2025.1206008

- **More Extensive Real-World Deployments / Field Trials**: Expand beyond simulation to actual test fleets across varied environments to validate system in practice.
- **Edge AI Model Compression & Robustness**: Further work on model compression, quantization, pruning, robust models that work under weather, lighting, sensor noise.
- **Safety, Explainability & Regulatory Compliance**: Develop explainable AI components; ensure systems satisfy automotive safety standards (e.g. ISO 26262 etc.); create legal and ethical frameworks for liability.
- **Infrastructure for V2X / 5G / Edge / Fog Integration**: Better deployment of RSUs, edge nodes; seamless communication (5G, low latency networks); orchestrating cloud/edge resources.
- **Energy Efficiency & Sustainability**: Minimize energy use in vehicles and edge infrastructure; explore green computing, renewable energy powering edge nodes; trade-offs between hardware complexity and energy.

REFERENCES

1. Sorin Grigorescu, Tiberiu Cocias, Bogdan Trasnea, Andrea Margheri, Federico Lombardi, Leonardo Aniello. Cloud2Edge Elastic AI Framework for Prototyping and Deployment of AI Inference Engines in Autonomous Vehicles. arXiv:2009.11722 (2020).
2. Sergey Andreev, Vitaly Petrov, Kaibin Huang, Maria A. Lema, Mischa Dohler. Dense Moving Fog for Intelligent IoT: Key Challenges and Opportunities. arXiv:1812.08387 (2018).
3. Dewant Katare, Diego Perino, Jari Nurmi, Martijn Warnier, Marijn Janssen, Aaron Yi Ding. A Survey on Approximate Edge AI for Energy Efficient Autonomous Driving Services. arXiv:2304.14271 (2023).
4. Advanced, Innovative AIoT and Edge–Cloud Computing for Unmanned Vehicle Systems in Factories. MDPI.
5. Khuram Shahzad, Cloud Robotics and Autonomous Vehicles. IntechOpen, (2016).
6. “Applications of AI, IoT, and Cloud Computing in Smart Transportation: A Review.” Mnyakin, M., et al., Artificial Intelligence in Society (2023).
7. Internet of Vehicles and its Applications in Autonomous Driving. SpringerLink (2021).
8. Intelligent Technologies for Internet of Vehicles. SpringerLink (2024).

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com