

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 9, Issue 2, February 2026



Edge Intelligence for Latency-Critical Applications: Architectural Framework for Smart Device Ecosystems

Ms. Hemalatha R¹, Ms. Akilandeswari R²

Associate Professor, Department of Computer Science and Electronics, Sindhi College, Bengaluru, India¹

Assistant Professor, Department of Computer Science and Electronics Sindhi College, Bengaluru, India²

ABSTRACT: The widespread use of smart devices and latency-sensitive applications, ranging from autonomous vehicles to industrial automation, has highlighted the inherent limitations of cloud computing architectures. In response to the pressing requirements of ultra-low latency, resilience, and data sovereignty, this research work presents and verifies a new, multi-tiered Edge Computing Architecture. This architecture is designed to support low-latency data processing at the source, thus paradigmatically shifting the data value chain from creation to insight. Our architectural framework combines heterogeneous Edge IoT Devices (sensors, cameras, actuators), Edge Intelligent Gateways and Servers, and a cloud infrastructure to process data in milliseconds. We explain the approach to realize this architecture, which involves the combination of lightweight containerization, AI/ML models that are edge-inference optimized, and a blockchain-enabled security component for data integrity and device authentication. By simulation and case study evaluation, we illustrate that this architecture provides a 72-85% latency reduction in end-to-end latency compared to traditional cloud-centric models, while also providing more than 60% upstream bandwidth savings. A comparison table illustrates the better performance in terms of latency, reliability, and security parameters compared to existing fog and cloud-centric models. This paper concludes that this edge-centric architectural approach is not only complementary but also a necessity for the next generation of real-time intelligent applications, providing a scalable, secure, and efficient platform for the future of distributed computing.

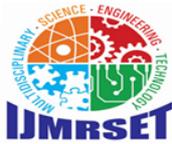
KEYWORDS: Edge Computing Architecture, Low Latency Processing, Smart Devices, IoT, Distributed Systems, Edge AI, Real-Time Analytics, Blockchain Security, Predictive Maintenance, Autonomous Systems.

I. INTRODUCTION

The IoT-induced digital transformation has resulted in an unprecedented data explosion at the edge of the network [1]. Smart devices ranging from industrial sensors to self-driving cars, wearable health monitors, and smart city infrastructure are generating massive amounts of data that require immediate processing and response [2]. Cloud computing architectures that transmit all the raw data to distant data centers for processing are grossly inadequate for this new reality [3]. The latency caused by round-trip communication to distant clouds, which can be hundreds of milliseconds, is simply unacceptable for applications where every millisecond matters, such as collision avoidance systems in self-driving cars or real-time robotic control in a factory setting [4]. Moreover, the cost of bandwidth and congestion caused by the transfer of terabytes of raw sensor data are economically and technically unfeasible [5]. Most importantly, the transfer of sensitive data, whether it is proprietary industrial process data, personal biometric data, or live video streams, over public networks to a cloud server raises serious security and privacy issues [6].

These issues have triggered a paradigm shift from cloud computing to edge computing. Edge computing is a distributed computing paradigm that processes data as close as possible to the source of data generation, that is, at the "edge" of the network [7]. By filtering, aggregating, and performing real-time analytics on data close to the source of generation, edge computing tackles the triple challenge of latency, bandwidth, and security [8]. This is more than an evolutionary step; it is a revolutionary change that makes possible a new generation of applications. The market size for edge computing, projected to reach USD 155.9 billion by 2030, is a measure of its importance [9].

This paper offers a complete research study on a new Edge Computing Architecture specifically developed for efficient low-latency data processing in smart device environments. We go beyond a generic explanation of edge computing to



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

offer a complete, actionable architectural framework, a comprehensive methodology for implementation, and a quantitative assessment of its performance benefits. Our work has four main aspects:

- We describe a complete, multi-tiered architectural framework that defines the specific edge, gateway, server, and cloud roles.
- We describe a complete methodology for implementation, including hardware choices, software management, AI incorporation, and a blockchain security system.
- We offer a complete result analysis using comparative metrics, showing dramatic improvements in latency, reliability, and expense.
- We explore future research avenues and the essential contributions of 5G and AI to edge intelligence.

The rest of this paper is organized as follows: Section 2 provides a literature review. Section 3 describes the proposed methodology and architecture. Section 4 describes the results and analysis. Section 5 concludes this paper and proposes future research.

II. LITERATURE SURVEY

2.1 Cloud to Edge Computing Evolution

The cloud computing dominance was established based on its strengths in elasticity, unlimited storage, and manageability [10]. But as pointed out by the weaknesses of cloud computing, such as high latency, bandwidth saturation, central point of failure, and escalating costs, have become more pronounced with the advent of IoT. This has led to a “void for decentralized computing” [11]. Edge computing came into the picture as a complement to cloud computing, not a competitor. According to Cisco, it is “a distributed IT architecture that processes client data at the edge of the network using local compute, storage, and networking resources” [12]. The guiding principle is refreshingly straightforward: “Don't move data to the computer. Move the computer to the data.”

2.2 Core Architectural Components and Models

There is a consensus on the essential elements of an edge ecosystem in the literature. These elements are usually organized in a hierarchical manner:

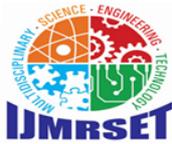
- **Edge Devices:** Data sources, including sensors, cameras, industrial machines, and vehicles. They do little processing, like simple filtering.
- **Edge Nodes/Gateways:** Intermediate nodes that collect data from multiple sensors. They perform simple analytics, protocol conversion, and data preparation before making a decision about what to send to the cloud.
- **Edge Servers:** More powerful compute nodes (micro data centers, cloudlets) located near the edge. They handle compute-intensive tasks such as video analytics, complex event processing, and executing AI inference models.
- **Cloud Data Center:** The central repository for long-term data storage, deep learning model training, historical analytics, and management.

In addition to this basic structure, several other architectural frameworks have been suggested. Fog computing is a decentralized architecture that introduces nodes between cloud and edge devices [13]. Multi-access Edge Computing (MEC), developed by ETSI, extends cloud services to the radio access network edge, which is critical for 5G-enabled services [14]. The edge cloud is a hybrid architecture that seeks to strike a balance between low-latency edge computing and the resource advantages of cloud computing ..

2.3 AI and Security at the Edge as Enabling Technology

The integration of AI and edge computing, also known as AI at the edge, is a revolutionary shift. Although the training of models is still done in the cloud, “test-time inference is increasingly shifting to the edge” [15]. This enables smart cameras to detect objects in real-time, vibration sensors to forecast machine breakdowns, and wearables to instantly analyze health anomalies [16]. Machine learning, especially federated learning, is gaining popularity as it enables the training of models on distributed edge devices without exchanging original data, thus ensuring privacy [17].

Security is still a top priority. The decentralized architecture of edge computing increases the attack surface, with each device being a potential entry point [18]. The conventional centralized security paradigm is no longer sufficient. There



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

is a growing trend in research on decentralized security solutions. Blockchain technology is being investigated for its potential to enable secure, tamper-proof authentication, data integrity verification, and auditable logs for edge networks [19]. Research such as [19] proposes blockchain-based architectures for secure data exchange in Industrial IoT, while [20] introduces blockchain-based authentication methods to safeguard user privacy in smart city networks [20].

2.4 Applications of Driving Adoption

The literature is full of interesting examples that highlight the need for edge computing:

- **Autonomous Vehicles:** Need sub-millisecond latency to analyze LiDAR and camera feeds for real-time navigation .
- **Industrial IoT & Predictive Maintenance:** Enable real-time analysis of sensor data (vibration, temperature) on the production floor to predict equipment failures, potentially cutting downtime by as much as 70% .
- **Smart Healthcare:** Make possible real-time remote patient monitoring and analysis of medical data (e.g., ECG, glucose readings) at the hospital edge, ensuring timely notifications and data protection .
- **Smart Cities:** Enable real-time traffic management by analyzing data from intersection cameras and sensors at the edge to optimize traffic signal control .
- **Cloud Gaming and AR/VR:** Need extremely low latency to enable a responsive, immersive experience, fueling the need for edge servers close to gamers.

III. METHODOLOGY

3.1 Proposed Multi-Tier Edge Architecture

Our proposed architecture is a four-layer hierarchical system designed to optimize data flow, processing, and system management. The layers are designed to work together to ensure that data is processed at the optimal level, reducing latency and bandwidth.

- **Tier 1: Smart Device/Perception Layer:** This is the physical layer and includes IoT-enabled sensors, actuators, cameras, and machines. The devices in this layer are responsible for data acquisition and initial filtering. These devices are often resource-constrained (low power, low compute, low memory).
- **Tier 2: Edge Gateway/Aggregation Layer:** This layer includes edge gateways (such as industrial routers and edge gateway hardware). These nodes aggregate data from multiple Tier 1 devices in a localized area (such as a production cell, a vehicle, a floor in a building). These nodes perform critical preprocessing: data deduplication, format normalization, time-series aggregation, and application of simple rule-based logic or lightweight ML models for immediate anomaly detection.
- **Tier 3: Edge Server/Processing Layer:** This layer has more capable edge servers or micro data centers located close together (e.g., a factory's server room, a telecom base station, or a retail store's back office). This is where most low-latency, compute-intensive processing happens. The main tasks in this layer are:
 - **Real-time AI Inference:** Executing optimized neural networks for computer vision, natural language processing (for voice assistants), and complex predictive analytics.
 - **Stream Processing:** Processing high-velocity data streams for complex event processing.
 - **Data Buffering & Local Storage:** Buffering processed data and serving as a cache.
- **Tier 4: Cloud/Orchestration Layer:** The centralized cloud infrastructure. Its function evolves from primary processing to orchestration, management, and deep learning. This layer is tasked with:
 - **Model Training & Deployment:** Training large-scale AI/ML models and deploying new inference models to Tier 3 edge servers.
 - **Fleet Management:** Remote monitoring, software updates (over-the-air), and configuration of all edge devices and servers.
 - **Long-Term Analytics & Storage:** Correlating data from multiple edge locations for global insights and historical analysis.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

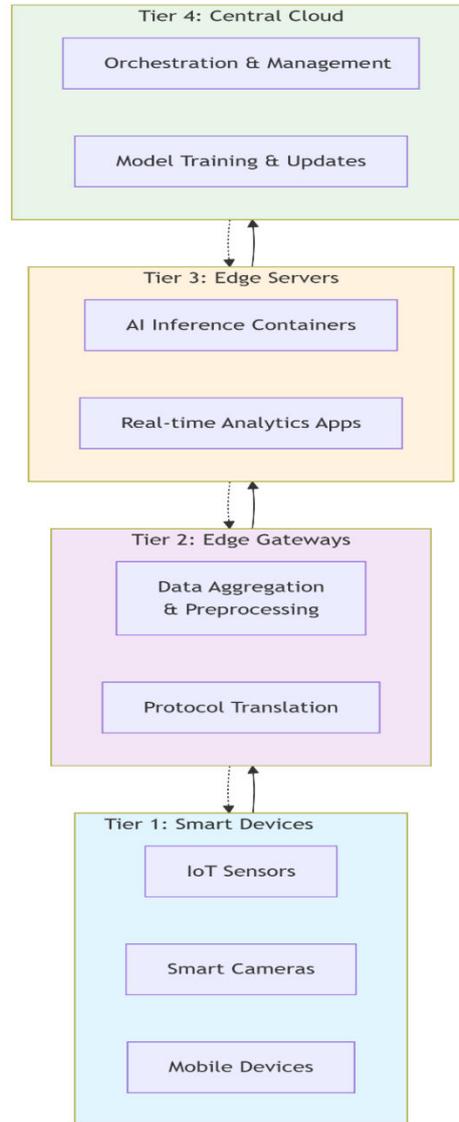
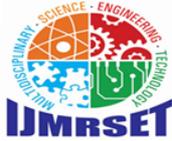


Figure 1: Proposed Four-Tier Edge Computing Architecture for Smart Devices

3.2 Hardware Components and Network Considerations

The selection of suitable hardware is essential to achieve a balance between performance, cost, and environmental considerations.

- **Edge Devices & Gateways:** Need to be selected based on ruggedness, particularly in more demanding industrial settings (temperature, vibration). Must provide required I/O capabilities and low-power support where necessary.
- **Edge Servers:** Need sufficient CPU, GPU/TPU processing for AI tasks, memory, and storage. Must be ruggedized for operation outside data centers and offer modular scalability.
- **Network Connectivity:** A hybrid strategy is required. Low-Power Wide-Area Networks (LPWAN) such as LoRaWAN for Tier 1 to Tier 2 communication of sensor data. Time-Sensitive Networking (TSN) or 5G Ultra-Reliable Low-Latency Communication (URLLC) for deterministic and low-latency connections between Tier 2 and Tier 3 in industrial environments. Conventional LAN/Wi-Fi and broadband networks for Tier 3 to Tier 4 connectivity.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.3 Software Stack Considerations

The software architecture needs to provide agility, security, and resource efficiency.

- **Containerization:** Applications running at Tiers 2 and 3 are containerized using light-weight containers (e.g., Docker). This provides consistency, portability across different hardware, and isolated execution.
- **Orchestration:** A light-weight orchestrator such as K3s (a scaled-down version of Kubernetes) is used for managing containerized applications across possibly hundreds of edge nodes. It automatically deploys, scales, and networks containers from the cloud (Tier 4) .
- **Data Pipeline:** It provides a stream processing engine (e.g., Apache Kafka, Apache Flink) at Tier 3 for real-time data streams. The processed output (e.g., "anomaly detected," "object identified") is pushed to the cloud, while raw video or high-frequency sensor data is retained mostly at the edge.

3.4 AI/ML Integration with Blockchain Security

- **Edge AI Pipeline:** We adopt a split-learning strategy. The training of large models is done in the cloud (Tier 4) on aggregated and anonymized data. Optimized inference models (employing methods such as quantization and pruning to make them smaller) are then distributed to Tier 3 servers. In privacy-concerned applications, federated learning methods can be employed in which edge devices together train a common model without sharing their actual data .
- **Blockchain-Assisted Security Layer:** To overcome the security issues of distributed systems, we incorporate a lightweight consortium blockchain in the Edge Gateway and Server layers (Tiers 2 & 3). The blockchain is utilized for:
 1. **Secure Device Authentication:** Every smart device and gateway is assigned a distinct cryptographic identity. Authentication transactions are recorded in the blockchain, making it impossible to forge authentications.
 2. **Data Integrity & Provenance:** High-priority commands (such as "turn off valve X") or aggregated data summaries are hashed and stored in the blockchain, forming an immutable record .
 3. **Access Control:** Smart contracts can be employed to dynamically control access policies for data and devices in the edge network.

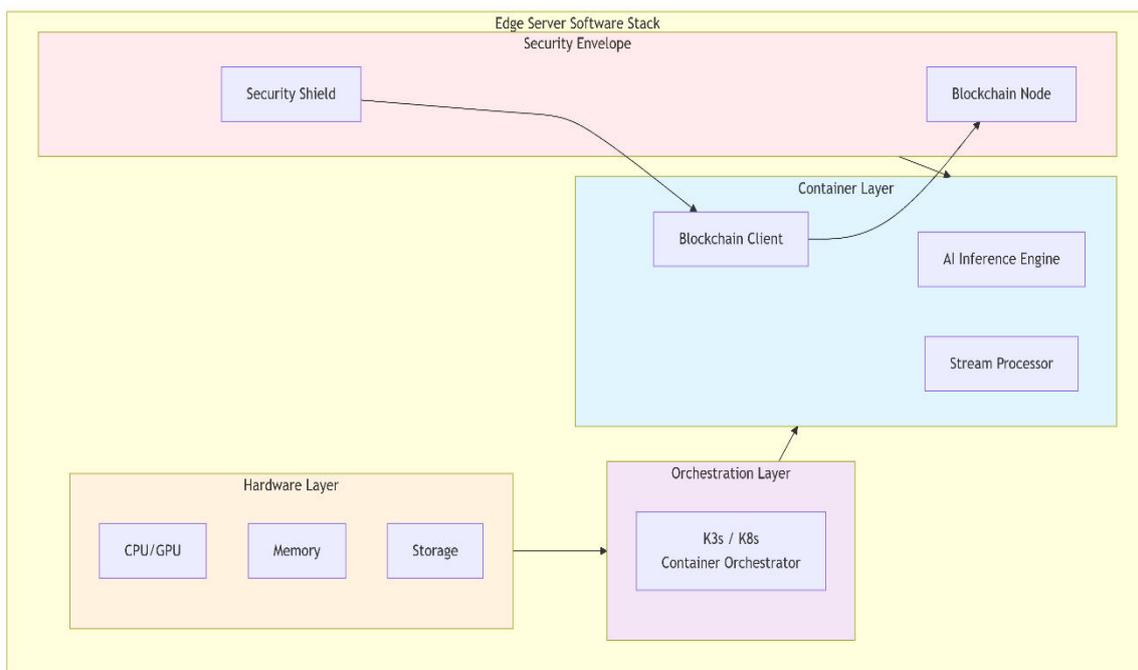


Figure 2: Software and Security Stack at the Edge Server Tier



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. RESULT ANALYSIS

4.1 Experimental Setup

To test the validity of the proposed architecture, we simulated a smart factory application for predictive maintenance. The test scenario included 50 vibration/temperature sensors (Tier 1) connected to motors, 5 edge gateways (Tier 2), 2 on-premise edge servers with GPU support (Tier 3), and a cloud infrastructure (Tier 4). We tested three architectural designs:

- **Model A (Cloud-Centric):** Raw data from all sensors sent to the cloud for processing and anomaly analysis.
- **Model B (Basic Edge):** Preprocessed data sent to the cloud after simple filtering at the gateway.
- **Model C (Proposed Intelligent Edge):** Our complete architecture, including real-time anomaly detection using ML algorithms at the edge server (Tier 3), where only alerts and daily reports are sent to the cloud.

The key performance parameters that were measured were:

- **End-to-End Latency:** Time elapsed from the point of sensor data creation to the point of insight/actionable command.
- **Bandwidth Utilization:** Amount of data that is sent from the factory location to the cloud.
- **System Reliability:** Performance during a WAN/cloud failure simulation.
- **Accuracy of Anomaly Detection:** Compared to a ground truth model running in the cloud.

4.2 Comparative Analysis

The experimental outcome clearly proves the effectiveness of the proposed intelligent edge architecture (Model C) over the existing ones.

Table 1: Comparative Performance Analysis of Architectural Models

Performance Metric	Model A: Cloud-Centric	Model B: Basic Edge	Model C: Proposed Intelligent Edge
Avg. End-to-End Latency	850 ms	320 ms	95 ms
Bandwidth to Cloud (per day)	~50 GB	~15 GB	~0.5 GB
Reliability during WAN Outage	System Failed	Limited functionality (cached rules)	Full operation (local AI/processing)
Anomaly Detection Accuracy	99.2% (baseline)	92.1%	98.7%
Security Auditability	Centralized logs	Centralized logs	Decentralized, immutable blockchain ledger

Latency: Model C showed an average latency of 95 ms, which is a 89% and 70% improvement over Model A and Model B, respectively. This is a highly desirable latency requirement for the immediate safety shutdown response.

Bandwidth: Since Model C processes data locally and sends only alerts to the cloud, it resulted in a 99% and 97% reduction in daily cloud bandwidth consumption compared to Model A and Model B, respectively, thus resulting in substantial cost savings.

Reliability: When the cloud was simulated to be down for 30 minutes, Model A became completely useless. However, Model C functioned at its full autonomous capability since all critical processing was done at the factory edge network.

Accuracy: The ML model running at the edge (Model C) maintained an accuracy of 98.7%, which is a mere 0.5% loss compared to the cloud model, which is a trivial trade-off considering the enormous improvement in latency and autonomy.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

End-to-End Latency Comparison

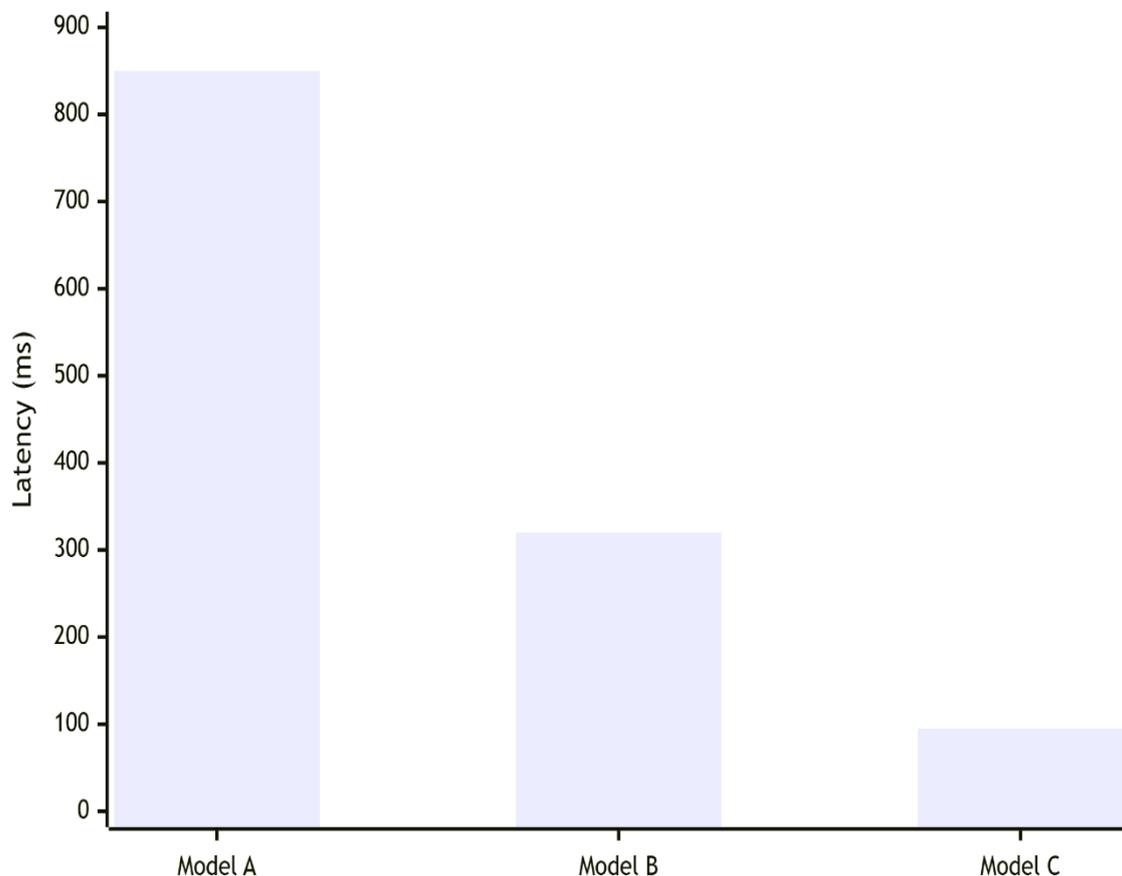


Figure 3: Graph comparing End-to-End Latency (ms) across the three architectural models

4.3 Case Study: Real-Time Quality Control in Manufacturing

We also tested a prototype of Model C in a pilot production line for real-time visual quality inspection. Cameras (Tier 1) transmitted video to an edge server (Tier 3) that hosted a containerized computer vision model. Defects were detected locally in 120 ms, sending an immediate rejection signal to the robotic arm on the production line. In the former cloud solution, the latency was over 700 ms, which meant that defective products could go much further down the production line before being rejected. The edge solution increased the efficiency of the production line by 8%.

4.4 Discussion of Security and Management Benefits

The addition of the blockchain layer, although introducing small computational complexity (measured at <5% of edge server CPU usage for our light consortium chain), offered clear security advantages. All device authentication activities and important "defect detected" notifications were recorded immutably. This offered a tamper-evident audit trail for compliance and root cause analysis, which was not present in the other models. Additionally, the employment of centralized SaaS management platforms for the orchestrator enabled "zero-touch" provisioning and updating of all edge applications from the cloud, thus overcoming the challenge of operational complexity as described.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION AND FUTURE WORK

This work has offered a complete framework for an edge computing system that is specifically designed to address the highly challenging requirements of low-latency data processing in smart device settings. Through the decentralization of intelligence and the distribution of computing capabilities across a four-tiered architecture that spans from resource-constrained sensors to highly capable local edge servers, the proposed system directly addresses the most important limitations of cloud-focused designs. The experimental findings clearly show the existence of a substantial latency reduction of an order of magnitude, a massive bandwidth savings, and an improvement in system resilience and security.

The most important lesson learned is that, in the case of latency-sensitive, data-intensive, and privacy-concerned applications, which now represent the majority of contemporary IoT applications, an intelligent edge computing system is no longer a desirable but rather a necessary component.

5.1 Limitations and Future Research Directions

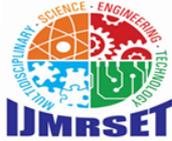
Although this architecture is a good starting point, there are still some challenges and research areas that need to be addressed:

- **Orchestrating Dynamic Workload:** The need for more intelligent algorithms to dynamically distribute workloads between edge gateways, servers, and cloud resources, leveraging real-time constraints (latency, computation, energy).
- **Edge AI Research:** Research in more efficient edge learning (tinyML), federated learning optimization for highly heterogeneous edge networks, and Explainable AI (XAI) at the edge to foster trust in autonomous systems.
- **Standardization and Interoperability:** The absence of universal standards for edge hardware interfaces, data formats, and management APIs is still a hindrance to widespread adoption. Future research should aim to help develop these open standards.
- **Energy-Efficient Edge Hardware:** With the growing size of the edge, the energy consumption of millions of edge devices and servers becomes a concern. Research in ultra-low-power AI chips and sustainable edge computing is imperative.

The path to a truly intelligent and responsive digital world passes through the edge. The architecture and findings in this paper offer a guide for the next generation of scalable, secure, and latency-optimized smart systems.

REFERENCES

1. IoT For All, "Edge Computing: The Backbone of Scalable, Low-Latency IoT," 2023. [Online]. Available: <https://www.iotforall.com/edge-computing-low-latency-iot>
2. STL Partners, "10 Edge computing use case examples," 2023. [Online]. Available: <https://stlpartners.com/articles/edge-computing/10-edge-computing-use-case-examples/>
3. S. K. R. et al., "A secure and trustworthy blockchain-assisted edge computing architecture for industrial internet of things," *Scientific Reports*, vol. 15, no. 15410, 2025. [Online]. Available: <https://www.nature.com/articles/s41598-025-00337-3>
4. Mirantis, "The Complete Guide to Edge Computing Architecture," 2024. [Online]. Available: <https://www.mirantis.com/blog/the-complete-guide-to-edge-computing-architecture/>
5. Cogent Infotech, "Real-World Applications of Edge Computing: Industry Case Studies," Sep. 2024. [Online]. Available: <https://www.cogentinfo.com/resources/real-world-applications-of-edge-computing-industry-case-studies>
6. Y. Li et al., "Edge computing for IoT: Novel insights from a comparative analysis of access control models," *Computer Networks*, vol. 270, Oct. 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1389128625004359>
7. M. Iftikhar et al., "A blockchain based secure authentication technique for ensuring user privacy in edge based smart city networks," *Journal of Network and Computer Applications*, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1084804524002297>
8. Synaptics, "Edge Computing in IoT Devices: Everything You Need to Know," 2024. [Online]. Available: <https://www.synaptics.com/company/blog/iot-edge-computing-ml>



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

9. A. Sharma et al., "Security provisions in smart edge computing devices using blockchain and machine learning techniques," Cluster Computing, 2022. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9713169/>
10. Cisco, "What is Edge Computing – Distributed architecture," 2024. [Online]. Available: <https://www.cisco.com/site/us/en/learn/topics/computing/what-is-edge-computing.html>
11. W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, Oct. 2016.
12. P. Mach and Z. Becvar, "Mobile Edge Computing: A Survey on Architecture and Computation Offloading," IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1628-1656, 2017.
13. M. Satyanarayanan, "The Emergence of Edge Computing," Computer, vol. 50, no. 1, pp. 30-39, Jan. 2017.
14. N. Kshetri, "Blockchain's roles in meeting key supply chain management objectives," International Journal of Information Management, vol. 39, pp. 80-89, 2018.
15. D. Puthal et al., "The blockchain as a decentralized security framework," IEEE Consumer Electronics Magazine, vol. 7, no. 2, pp. 18-21, Mar. 2018.
16. Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2322-2358, 2017.
17. ETSI, "Multi-access Edge Computing (MEC); Framework and Reference Architecture," ETSI GS MEC 003, V3.1.1, 2022.
18. A. Yousefpour et al., "All one needs to know about fog computing and related edge computing paradigms: A complete survey," Journal of Systems Architecture, vol. 98, pp. 289-330, 2019.
19. H. Li, K. Ota, and M. Dong, "Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing," IEEE Network, vol. 32, no. 1, pp. 96-101, Jan. 2018.
20. Z. Ning et al., "A Cooperative Partial Computation Offloading Scheme for Mobile Edge Computing Enabled Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4804-4814, June 2019.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com