Volume 01, Issue 01, January - February 2026
ISSN (Print) : 3139 - 471X
ISSN (Online): ****_****

Journal of Emerging Technologies and Innovative Research Development (JETIRD)

https://jetird.com/

Hybrid AI Models for Real-Time lot Data Analytics on Edge-Cloud Platforms

[1*]Dr. S. B. Sawant, [2]Koppula Sridhar

[1]Director, Department of MBA, Abhijit Kadam Institute of Management and Social Sciences, Bharati Vidyapeeth Deemed to be University, Solapur, Maharashtra.

[2]Assistant Professor, Department Computer Science And Engineering, CMR Engineering College. Kondlakoya, Medchal

ABSTRACT

The rapid growth of the Internet of Things (IoT) has led to massive volumes of distributed, high-velocity data that require low-latency, privacy-preserving, and scalable analytics. Pure cloud-based processing often fails to meet real-time constraints, while pure edge analytics struggle with limited computation and storage resources. Hybrid AI, which combines edge and cloud intelligence, has emerged as a promising paradigm for real-time IoT data analytics. This paper proposes a conceptual framework for hybrid AI models on edge–cloud platforms, where latency-sensitive inference is executed at the edge while model training, global optimization, and cross-domain analytics are offloaded to the cloud. We review recent advances in edge and hybrid architectures, discuss AI deployment patterns, and present a reference architecture with a layered data flow. A comparative table highlights key trade-offs between cloud-only, edge-only, and hybrid AI strategies. The paper concludes with open challenges in orchestration, resource management, security, and standardization.

KEYWORDS

hybrid AI, edge computing, cloud computing, IoT, real-time analytics, federated learning

ARTICLE HISTORY

Received: 16 January 2026
Accepted: 05 February 2026
Published: 17 February 2026

CITATION

Sawant, S. B., Hybrid AI Models for Real-Time lot Data Analytics on Edge-Cloud Platforms. *Journal of Emerging Technologies and Innovative Research Development (JETIRD*, 1(1), 49-54.
Doi

Introduction

IoT deployments in domains such as smart factories, healthcare, transportation, and precision agriculture generate continuous streams of sensor data that must be processed in (near) real time to enable timely decisions and automated control. Traditional cloud-centric architectures centralize data storage and computation in remote data centers. While this offers elastic compute and powerful AI capabilities, it introduces latency, bandwidth costs, and potential privacy issues—particularly for time-sensitive or sensitive applications. Edge computing complements the cloud by pushing computation closer to data sources at gateways, routers, and micro data centers, thereby reducing round-trip delay and network congestion. However, edge nodes are resource-constrained and cannot alone support complex AI pipelines or long-term storage. To address these limitations, hybrid AI architectures combine edge AI and cloud AI: models are trained or refined in the cloud, deployed for inference at the edge, and continuously updated based on feedback and local learning. Recent studies in AIoT systems and hybrid edge–cloud frameworks highlight the need for such integrated approaches to satisfy both real-time and global optimization requirements. This paper focuses on Hybrid AI Models for Real-Time IoT Data Analytics on Edge–Cloud Platforms and makes three contributions:

1. Synthesizes recent literature on edge, cloud, and hybrid architectures for IoT analytics.
2. Proposes a layered hybrid AI reference architecture for real-time analytics.
3. Provides a comparative analysis of cloud-only, edge-only, and hybrid AI deployment models and discusses open challenges.

2. Background and Related Work

2.1 Edge and Cloud Computing in IoT

Edge computing is characterized by deploying compute and storage resources near IoT devices to minimize latency and reduce backbone traffic. Surveys on edge computing for IoT report significant benefits in response time, bandwidth savings, and energy efficiency, especially for real-time applications. Cloud computing remains essential to provide large-scale storage, powerful accelerators (GPU/TPU), and centralized management. Recent reviews emphasize that hybrid edge–cloud architectures are preferable for complex IoT workloads, combining low latency at the edge with elasticity and global visibility in the cloud.

2.2 AI in IoT Systems and the Case for Hybrid Deployment

A recent systematic literature review on AI in IoT systems finds that most deployments are either cloud-centric or device-centric, with relatively limited exploration of dynamic, hybrid deployment strategies that balance computation between edge and

Industrial platforms such as Sophon Edge demonstrate edge–cloud collaborative computing for AIoT applications, using pipeline-based models for streaming data and iterative model evolution in the cloud. Application-specific systems, such as precision livestock monitoring, have successfully integrated hybrid AI (gradient boosting + LSTM) with IoT, edge, and cloud computing for real-time prediction and control. These studies collectively underscore that hybrid AI is emerging as a practical solution for real-time analytics in complex IoT environments, but systematic architectural guidance is still evolving.

2.3 Federated and Collaborative Learning

Federated learning (FL) enables distributed model training across edge devices without centralizing raw data, reducing bandwidth usage and improving privacy—making it well suited for IoT and edge scenarios. FL can be combined with edge–cloud collaborations, where local models are trained at the edge and aggregated in the cloud, further reinforcing the hybrid AI paradigm.

3. Hybrid AI for Edge–Cloud IoT Analytics

In this paper, Hybrid AI refers to the combined use of:

- Edge AI for on-device or near-device inference on streaming data,
- Cloud AI for centralized training, model orchestration, and cross-domain analytics, and
- Collaborative learning mechanisms (e.g., FL, model distillation, continual learning) to synchronize models.

Hybrid AI seeks to:

- Minimize end-to-end latency for critical decisions,
- Reduce raw data transmission to the cloud,
- Maintain or improve model accuracy via cloud-based training and global context, and
- Support dynamic workload offloading based on current network and resource conditions.tierpoint.com+2ekascloud.com+2

4. Proposed Reference Architecture

4.1 Architectural Overview

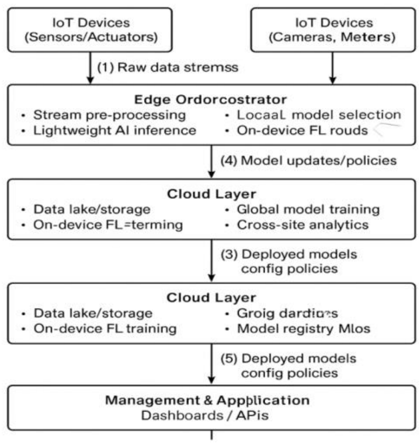

Figure 1. Conceptual hybrid AI architecture for edge cloud IoT analytics

Data flow (high level):

1. **Data generation:** IoT devices produce continuous data streams.
2. **Edge analytics:** Edge nodes perform pre-processing, filtering, and low-latency inference (e.g., anomaly detection, threshold alerts). Only summaries, selected events, or compressed features are forwarded.
3. **Cloud analytics:** The cloud aggregates edge outputs, performs batch analytics, model training, and global optimization using historical and cross-domain data.
4. **Model coordination:** Updated models, hyperparameters, or policies are distributed back to the edge; optionally, FL or other collaborative learning protocols are used.
5. **Applications:** Dashboards, APIs, and control systems consume both real-time edge insights and cloud-derived analytics.

4.2 Hybrid AI Model Patterns

Typical hybrid AI deployment patterns include:

- **Cloud-trained, edge-deployed models:** heavy training in the cloud; compressed models (e.g., quantized, pruned) deployed to gateways or devices for inference.
- **Edge-preprocessing + cloud inference:** simple feature extraction at edge, inference in cloud for moderately latency-sensitive tasks.

Volume 01, Issue 01, January - February 2026
ISSN (Print) : 3139 - 471X
ISSN (Online): ****_****

- **Federated edge training + cloud aggregation:** local models trained at edge; cloud aggregates weights (FL) and redistributes improved models.

5. Comparative Analysis of Deployment Strategies

Table 1 compares cloud-only, edge-only, and hybrid AI strategies for real-time IoT analytics.

Table 1. Comparison of AI deployment strategies for IoT analytics

Approach	Latency (to decision)	Bandwidth Usage	Model Complexity & Training	Privacy & Data Locality	Typical Use Cases
Cloud-only AI	Medium to high (depends on network RTT)	High – large volumes of raw/semiprocessed data sent to cloud	High – complex deep models easily supported	Low – raw data often centralized	Non-real-time analytics, batch reporting, global optimization, long-term forecasting
Edge-only AI	Very low – decisions near source	Low – minimal upstream data	Limited by edge resources; retraining difficult	High – data stays local	Ultra-low-latency control, offline operations, privacy-critical domains with small models
Hybrid Edge–Cloud AI	Low for local decisions; medium for cloud-enhanced tasks	Moderate – features, summaries, model updates instead of full raw streams	High – cloud supports heavy training; edge hosts optimized inference models	High to medium – raw data often local; only derived information shared	Real-time monitoring with global optimization (smart factories, smart cities, precision agriculture, e-health)

This comparison aligns with recent reviews that show hybrid architectures can balance the trade-offs between latency, bandwidth, and analytics sophistication in IoT environments.MDPI+1

6. Example Hybrid AI Workflow (Conceptual)

To illustrate how the architecture operates, consider a smart manufacturing scenario with hundreds of sensors monitoring vibration, temperature, and power consumption in production lines.

1. **Edge-level processing:**

- Edge nodes run lightweight anomaly detection (e.g., online clustering or one-class models) on vibration data for each machine.
- When anomalies exceed thresholds, the edge generates local alerts and may trigger immediate actions (e.g., slow-down or stop a machine).

2. **Cloud-level analytics:**

- Edge nodes periodically send compressed feature vectors and anomaly summaries to the cloud.

- Cloud trains more complex hybrid models (e.g., CNN–LSTM or gradient boosting + temporal networks) to predict remaining useful life (RUL) and cross-machine patterns.

3. Model distribution:

- The cloud updates edge models weekly based on global data (e.g., improved thresholds, updated anomaly classifier).
- For privacy, only model parameters are sent; raw sensor traces stay local. FL may be used where factories are geographically distributed and data cannot leave the premises.

4. Application layer:

- Dashboards show real-time alarms from edge plus cloud-derived RUL estimates, enabling predictive maintenance scheduling and reduced downtime.

7. Challenges and Research Directions

Despite the promise of hybrid AI for edge–cloud IoT analytics, several open challenges remain:

7.1 Dynamic Orchestration and Resource Management

Deciding **where** to execute which part of an AI pipeline (edge vs. cloud) in real time depends on network conditions, resource availability, and application priorities. Recent work on edge–cloud collaborative auto-scaling shows significant gains in latency and bandwidth, but generalizable orchestration policies remain an active research area.

7.2 Model Lifecycle and MLOps Across Edge and Cloud

Managing model versions, rollbacks, and A/B testing across thousands of distributed edge nodes is non-trivial. AIoT platforms like Sophon Edge and hybrid industrial solutions demonstrate pipeline-based model evolution, but standardized MLOps practices tailored to edge–cloud environments are still emerging.

7.3 Security, Privacy, and Trust

Edge devices are often physically exposed and vulnerable to tampering. Surveys on edge security highlight challenges including secure boot, attestation, and key management.IJCRT Hybrid AI must integrate secure communication, encrypted model updates, and robust mechanisms against model poisoning or adversarial attacks, especially when FL is employed.

7.4 Standardization and Interoperability

Heterogeneous hardware, operating systems, and vendor platforms make it difficult to deploy uniform hybrid AI solutions. Reviews emphasize the need for standardized APIs, containerization, and portable model formats (e.g., ONNX) to facilitate interoperability across edge and cloud environments.

7.5 Evaluation Benchmarks

Systematic evaluations of hybrid AI strategies across different application domains (smart cities, healthcare, industry 4.0) are still limited. Recent surveys call for benchmark suites that measure not only accuracy but also latency, energy, bandwidth, and robustness under realistic IoT workloads.

8. Conclusion

Hybrid AI models deployed on edge–cloud platforms represent a promising direction for real-time IoT data analytics. By combining edge-level, low-latency inference with cloud-based training and global analytics, they can meet stringent timing and privacy requirements while leveraging the full power of modern AI.

This paper has:

- Reviewed key literature on edge, cloud, and hybrid computing for IoT;
- Proposed a conceptual hybrid AI architecture with a clear data flow; and
- Compared cloud-only, edge-only, and hybrid AI strategies using an analytic table.

Future work should focus on dynamic orchestration mechanisms, comprehensive MLOps for distributed AI, robust security for model and data flows, and standardized benchmarks that capture end-to-end system performance.

References

1. Andriulo, F. C., Fiore, M., Mongiello, M., Traversa, E., & Zizzo, V. (2024). Edge computing and cloud computing for Internet of Things: A review. *Informatics, 11*(4), 71. https://doi.org/10.3390/informatics11040071
2. Desai, R., & co-authors. (2025). E-Cloud collaborative framework for real-time IoT data processing. *Journal of Cloud Computing and Information Technology*.admin.mantechpublications.com
3. Jebari, H., Rekiek, S., & Reklaoui, K. (2025). Advancing precision livestock farming: Integrating hybrid AI, IoT, cloud and edge computing for enhanced welfare and efficiency. *International Journal of Advanced Computer Science and Applications*, *16*(7), Article 32.
4. Khadam, U., Davidsson, P., & Spalazzese, R. (2025). A systematic literature review on AI in IoT systems: Tasks, applications, and deployment. *Internet of Things, 34*, 101779.ScienceDirect
5. Kong, L., Tan, J., Huang, J., Chen, G., Wang, S., Jin, X., Zeng, P., Khan, M. K., & Das, S. K. (2022). Edge-computing-driven Internet of Things: A survey. *ACM Computing Surveys, 55*(8), Article 174.
6. Mohammed, M. A., et al. (2025). Federated learning-driven IoT and edge cloud networks for healthcare monitoring. *International Journal of Computer Science and Mobile Computing*, *6*(1), 103–118.ijcsm.researchcommons.org+1
7. Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila, M., & Taleb, T. (2018). Survey on multi-access edge computing for Internet of Things realization. *IEEE Communications Surveys & Tutorials*, *20*(4), 2961–2991.ACM Digital Library
8. Rong, G., et al. (2021). An edge-cloud collaborative computing platform for building AIoT applications. *Journal of Cloud Computing*, 10, Article 40.SpringerOpen
9. Sheikh, A. M., et al. (2025). A survey on edge computing security challenges. *Future Internet*, *17*(4), 175.
10. Zhou, J., et al. (2024). Enhancing quality of service through federated learning in edge–cloud AIoT systems. *Computer Networks*, *244*, 110652.ScienceDirect+1