

Edge Computing Frameworks for Low-Latency Industrial Decision-Making

Enzo Mingozzi

Professor, University of Pisa, Italy

Abstract

The increasing adoption of connected industrial systems has created significant demand for computational architectures capable of processing data with minimal latency. Conventional cloud-centric computing models may introduce communication delays, bandwidth limitations, and dependence on remote data centers when large volumes of industrial information must be processed in real time. Edge computing addresses these limitations by placing computational resources closer to industrial devices, machines, sensors, and control systems. This paper examines edge computing frameworks for low-latency industrial decision-making, focusing on architectural components, data processing mechanisms, resource management, intelligent analytics, and industrial applications. A structured review-based methodology is adopted to analyze edge-enabled approaches across manufacturing, industrial automation, energy systems, transportation, robotics, and infrastructure.

The analysis demonstrates that edge computing can reduce response time by enabling localized data processing and decision-making while reducing unnecessary data transmission to centralized cloud environments. Integration with artificial intelligence, machine learning, Internet of Things technologies, and containerized services can further improve the intelligence and flexibility of edge platforms. However, practical implementation remains affected by limited edge resources, heterogeneous devices, network reliability, cybersecurity, interoperability, scalability, and workload management. The study highlights the importance of hierarchical edge-cloud architectures that distribute computation according to latency, resource, security, and application requirements. Future research should focus on adaptive resource orchestration, lightweight AI models, federated intelligence, secure edge architectures, digital twins, and autonomous industrial decision-making. Edge computing therefore represents a critical architectural approach for developing responsive, intelligent, and reliable industrial systems where rapid decisions are essential.

Keywords: Edge Computing, Industrial Decision-Making, Low Latency, Industrial IoT, Edge Intelligence, Cloud Computing, Real-Time Processing, Smart Manufacturing.

1. Introduction

The transformation of industrial systems into highly connected and intelligent environments has substantially increased the volume and velocity of data generated by machines, sensors, robots, production equipment, and control systems. Industrial organizations increasingly require rapid analysis of this information to monitor operations, identify abnormalities, optimize processes, and respond to changing conditions. Traditional cloud computing architectures provide substantial computational and storage resources, but transmitting every data stream to centralized cloud infrastructure can introduce

communication delays and bandwidth requirements. Edge computing has emerged as an architectural approach for addressing these limitations by bringing computational resources closer to data sources.

Edge computing distributes processing capabilities across locations near industrial devices rather than relying exclusively on centralized cloud platforms. This arrangement enables industrial data to be analyzed locally, allowing time-sensitive decisions to be made without waiting for communication with remote data centers. The concept is particularly relevant for applications where milliseconds or seconds can influence operational outcomes, including industrial automation, robotic control, machine monitoring, autonomous transportation, and safety systems.

The development of Industrial Internet of Things (IIoT) technologies has further increased the importance of edge computing. Industrial sensors can generate continuous streams of information related to temperature, vibration, pressure, speed, energy consumption, production quality, and machine condition. Sending all raw data to centralized infrastructure can create network congestion and unnecessary communication overhead. Edge platforms can filter, aggregate, analyze, and respond to information locally before forwarding only relevant data to cloud environments.

Edge computing can also support artificial intelligence and machine learning at the industrial level. Instead of transmitting all operational information to a remote server for inference, trained models can be deployed directly on edge devices or gateways. This enables rapid anomaly detection, predictive maintenance, quality inspection, and process optimization. The emerging concept of edge intelligence therefore combines localized computation with AI capabilities to support faster industrial decisions.

Nevertheless, industrial edge environments are more constrained than centralized cloud platforms. Edge devices may have limited processing power, memory, storage, and energy resources. Industrial networks also contain heterogeneous devices using different protocols and communication standards. Managing workloads across these distributed environments is therefore considerably more complex than processing data within a centralized infrastructure.

Security is another important concern. Edge devices may be physically distributed across factories, transportation networks, energy facilities, and other operational environments. Compromised edge devices could affect both information systems and physical processes. Consequently, secure authentication, encrypted communication, access management, trusted execution, and continuous monitoring are essential components of industrial edge architectures.

This paper examines edge computing frameworks for low-latency industrial decision-making. It analyzes edge architecture, data processing, intelligent analytics, resource management, industrial applications, implementation challenges, and future research directions.

2. Objectives

The primary objective of this study is to examine edge computing frameworks capable of supporting low-latency decision-making in industrial environments. The study focuses on how distributed computational resources can process industrial data close to its source while reducing communication delays, network congestion, and dependence on centralized cloud infrastructure.

2.1 Evaluation of Edge Computing Frameworks

The first objective is to examine major architectural approaches for industrial edge computing, including device-edge, edge-cloud, hierarchical, and distributed edge architectures. The study evaluates their

ability to process industrial data, support real-time analytics, manage computational workloads, and satisfy application-specific latency requirements.

The second objective is to analyze the role of edge intelligence in industrial decision-making. Particular attention is given to machine learning inference, real-time analytics, data filtering, event detection, and localized control. The study considers how these capabilities can improve response time and operational efficiency while reducing unnecessary communication with cloud platforms.

2.2 Assessment of Industrial Applications

The study further aims to assess edge computing applications across smart manufacturing, industrial automation, robotics, energy, transportation, and infrastructure. It examines how localized processing can support predictive maintenance, quality control, process monitoring, autonomous decision-making, and safety-related operations.

An additional objective is to identify the technical challenges associated with industrial edge deployment. These include resource constraints, heterogeneous devices, network reliability, cybersecurity, scalability, interoperability, workload scheduling, and distributed system management.

3. Review of Literature

3.1 Edge Computing and Distributed Industrial Systems

Edge computing represents a shift from centralized processing toward distributed computation closer to data-generation points. Shi et al. described edge computing as an approach that moves computation and services toward the network edge to address latency, bandwidth, and location-awareness requirements. This concept has become increasingly relevant to industrial systems where rapid responses are required.

Satyanarayanan highlighted the importance of edge computing for applications requiring low latency and high responsiveness. Industrial systems represent an important use case because many decisions must be made near physical processes rather than after data have traveled to distant cloud environments.

3.2 Industrial Internet of Things and Edge Computing

The Industrial Internet of Things connects machines, sensors, controllers, and industrial infrastructure to digital networks. This connectivity produces large quantities of operational information that can be analyzed for monitoring and optimization. Edge computing provides an intermediate processing layer between industrial devices and centralized cloud services.

Edge-based IIoT architectures can reduce the amount of raw information transmitted through networks. Local processing can identify relevant events and forward summarized or selected information to centralized systems. This can improve bandwidth utilization while maintaining access to cloud resources for long-term analytics and storage.

3.3 Edge Intelligence and Machine Learning

The integration of machine learning with edge computing has resulted in the development of edge intelligence. Machine learning models can be deployed on edge devices to perform inference locally. This is useful for applications such as visual inspection, anomaly detection, predictive maintenance, and equipment monitoring.

However, deploying machine learning models at the edge introduces computational constraints. Deep learning models can be resource intensive, requiring model compression, quantization, pruning, or lightweight architectures. Effective edge intelligence therefore requires balancing predictive performance against resource consumption and response-time requirements.

3.4 Edge-Cloud Collaboration

Edge and cloud computing are increasingly considered complementary components of distributed computing architectures. Cloud infrastructure provides large-scale processing, storage, model training, and historical analytics, whereas edge infrastructure provides rapid local processing and immediate responses.

An industrial edge-cloud framework can therefore divide workloads according to application requirements. Latency-sensitive tasks can be executed at the edge, while computationally intensive or long-term analytical tasks can be transferred to cloud infrastructure. This hierarchical approach can improve overall system efficiency.

4. Methodology

4.1 Research Design

This study uses a structured review-based methodology to examine edge computing frameworks for low-latency industrial decision-making. Literature was conceptually analyzed according to architecture, computing location, latency requirements, intelligent processing capabilities, industrial application, and implementation challenges. Publications available through 2024 were considered in accordance with the defined reference-period requirement.

The reviewed approaches were classified into device-level edge, gateway-level edge, distributed edge, hierarchical edge-cloud, and intelligent edge architectures. Industrial applications were organized into manufacturing, automation, robotics, energy, transportation, and infrastructure. This classification provides a structured basis for comparing architectural characteristics and industrial suitability.

4.2 Analytical Framework

The analytical framework evaluates edge computing frameworks according to latency reduction, computational efficiency, bandwidth utilization, scalability, resource management, reliability, security, and interoperability. Latency is treated as a central criterion because industrial decision-making often requires rapid responses to changing physical conditions.

The framework also considers the relationship between edge and cloud resources. An efficient industrial architecture should determine which tasks should remain local and which can be transferred to centralized infrastructure. This requires consideration of workload complexity, data sensitivity, response-time requirements, computational capacity, and network conditions.

4.3 Key Evaluation Parameters

The principal evaluation parameters include:

- **Latency:** time required to process data and produce a decision.
- **Bandwidth efficiency:** reduction in unnecessary data transmission.
- **Computational efficiency:** effective use of limited edge resources.
- **Scalability:** ability to support increasing numbers of devices.

- **Reliability:** continuity of industrial decision-making.
- **Security:** protection of distributed industrial resources.
- **Interoperability:** compatibility among heterogeneous devices and protocols.
- **Adaptability:** ability to adjust resource allocation according to workload changes.

5. Analysis

5.1 Edge Computing Architectural Models

Industrial edge computing can be implemented through several architectural models. Device-level edge processing performs computation directly on sensors, machines, controllers, or embedded devices. Gateway-level processing uses intermediate edge gateways that collect information from multiple devices. Distributed architectures use multiple edge nodes to share workloads, while hierarchical architectures connect device, edge, fog, and cloud layers.

Table 1: Comparative Characteristics of Industrial Edge Computing Frameworks

Framework	Processing Location	Latency Potential	Major Strength	Main Limitation
Device Edge	Sensors/industrial devices	Very low	Immediate processing	Limited resources
Gateway Edge	Local gateways	Low	Aggregates multiple devices	Gateway dependency
Distributed Edge	Multiple edge nodes	Low	Workload distribution	Coordination complexity
Hierarchical Edge-Cloud	Edge + cloud	Low to moderate	Flexible resource allocation	Management complexity
AI Edge	Edge devices/nodes	Very low	Local intelligent inference	Model resource requirements
Hybrid Edge-Fog-Cloud	Multiple distributed layers	Low	Broad computational flexibility	High orchestration complexity

The comparison indicates that device-level and AI-enabled edge architectures are particularly suitable for extremely latency-sensitive applications. However, their limited resources can restrict the complexity of computational workloads. Hierarchical edge-cloud architectures provide greater flexibility because they allow tasks to be distributed according to application requirements.

5.2 Industrial Applications

Edge computing has substantial potential across industrial environments. Manufacturing systems can use edge nodes for real-time quality inspection, equipment monitoring, machine coordination, and production optimization. Robotics applications can process sensor and vision information locally to

reduce response times. Predictive maintenance systems can analyze machine data at the edge and identify abnormal behavior before transmitting detailed information to cloud infrastructure.

Energy systems can use edge computing for local monitoring and rapid control of distributed energy assets. Transportation applications can process vehicle and infrastructure information close to the point of collection. Infrastructure systems can similarly use edge nodes for monitoring structures, equipment, and environmental conditions.

Table 2: Industrial Applications of Low-Latency Edge Computing

Industrial Domain	Application	Edge Processing Function	Expected Benefit
Smart Manufacturing	Quality inspection	Local image/data analysis	Faster defect detection
Industrial Automation	Process control	Real-time local decisions	Reduced response time
Robotics	Motion and environment analysis	Local sensor processing	Faster control
Predictive Maintenance	Fault detection	Local anomaly analysis	Earlier intervention
Energy Systems	Distributed monitoring	Local analytics and control	Improved responsiveness
Transportation	Intelligent traffic systems	Local event processing	Faster decisions
Infrastructure	Asset monitoring	Edge-based condition analysis	Rapid fault identification
Industrial Safety	Hazard detection	Immediate event processing	Faster safety response

The application analysis demonstrates that edge computing is most valuable when decisions must be made rapidly and where continuous transmission of raw data to centralized systems would introduce unnecessary latency or bandwidth consumption.

5.3 Edge-Based Intelligent Decision-Making

Low-latency decision-making involves more than simply moving computation closer to industrial devices. Edge platforms must acquire data, preprocess information, execute analytical models, determine appropriate actions, and communicate decisions to industrial control systems. The complete process must occur within application-specific response-time requirements. Machine learning can strengthen this process by allowing edge nodes to detect patterns and classify events locally. For example, an edge node monitoring industrial machinery can identify abnormal vibration patterns and immediately issue a

maintenance alert. Only selected information may then be transferred to cloud systems for long-term analysis.

The effectiveness of edge intelligence therefore depends on both computational placement and model efficiency. Lightweight machine learning models, model compression, hardware acceleration, and optimized inference pipelines can help ensure that intelligent processing remains within edge resource limitations.

6. Discussion

6.1 Edge Computing for Real-Time Industrial Decisions

The analysis indicates that edge computing can substantially improve industrial responsiveness by reducing the physical and computational distance between data sources and decision mechanisms. This is especially important for control systems, robotics, safety applications, and automated manufacturing processes where delayed responses may reduce efficiency or create operational risks.

Local processing can also improve system resilience. If communication with centralized cloud infrastructure is temporarily interrupted, edge systems can continue performing critical local operations. This makes edge computing particularly relevant for industrial environments requiring high availability.

6.2 Integration of Edge Computing and Artificial Intelligence

AI integration transforms edge infrastructure from a simple data-processing layer into an intelligent decision-making platform. Edge-based AI can analyze sensor information, identify anomalies, classify events, and generate recommendations without continuously transmitting raw data to centralized servers. However, AI models must be adapted to edge environments.

Large models may exceed local computational capacity or create unacceptable inference delays. Model compression, pruning, quantization, efficient architectures, and specialized hardware can therefore be important for deploying intelligent models at the industrial edge.

6.3 Edge-Cloud Collaboration

The analysis supports a complementary relationship between edge and cloud computing. Edge systems are well suited to immediate decisions, while cloud infrastructure remains valuable for large-scale historical analysis, centralized model training, storage, and strategic optimization.

An effective architecture should therefore dynamically allocate workloads. Tasks requiring immediate responses should remain close to industrial devices, while computationally intensive tasks without strict latency requirements can be transferred to cloud infrastructure. This workload distribution can improve resource utilization while maintaining industrial responsiveness.

7. Challenges

A major challenge in industrial edge computing is limited computational capacity. Edge devices generally have fewer processing, memory, and storage resources than centralized cloud systems. Industrial applications may nevertheless require sophisticated analytics, machine learning, and real-time processing. Efficient workload management and lightweight algorithms are therefore essential.

Heterogeneity is another significant challenge. Industrial environments contain devices from different manufacturers, generations, and communication standards. Integrating these devices into a

unified edge framework can create interoperability difficulties. Standardized communication interfaces and middleware are required to simplify integration.

Cybersecurity is particularly important because edge nodes are distributed across physical industrial environments. These nodes may be exposed to unauthorized access, physical tampering, malicious software, and network attacks. Security mechanisms must therefore be implemented at device, communication, application, and management levels.

Network reliability also affects edge-cloud collaboration. Although edge computing reduces dependence on continuous cloud communication, many applications still require periodic synchronization. Network disruptions, congestion, or unstable connections can interfere with workload migration and data synchronization.

Scalability represents another challenge. Industrial deployments can involve thousands of sensors and numerous edge nodes. Managing computational resources, software updates, AI models, and security policies across such distributed environments becomes increasingly difficult as system size increases.

8. Future Research

8.1 Adaptive Edge Resource Optimization

Future research should develop adaptive resource-management techniques capable of dynamically allocating processing, memory, and communication resources according to industrial workload requirements. AI-based orchestration can predict workload changes and automatically determine whether tasks should be executed locally or transferred to other edge nodes or cloud platforms.

Such approaches can improve resource utilization while maintaining latency requirements. Future systems should also incorporate energy efficiency because many edge devices operate under limited power budgets.

8.2 Federated and Privacy-Preserving Edge Intelligence

Federated learning provides a promising direction for industrial edge intelligence because models can be trained across distributed devices without continuously transferring raw operational data to a central server. This approach can improve privacy while allowing multiple industrial locations to contribute to shared models.

Future research should investigate secure federated learning mechanisms designed specifically for heterogeneous industrial environments. Attention should be given to communication efficiency, model synchronization, malicious participants, and variations in local datasets.

8.3 Autonomous Edge-Cloud Industrial Systems

Future industrial frameworks are likely to become increasingly autonomous. Edge nodes could continuously monitor industrial processes, predict workload requirements, detect abnormal conditions, and automatically optimize computational placement. Digital twins can further provide virtual representations of industrial processes for evaluating optimization decisions before deployment.

The combination of edge computing, AI, digital twins, and autonomous orchestration could create industrial systems capable of adapting computational resources according to changing physical and operational conditions while maintaining predefined safety and performance constraints.

9. Limitations

This study is based on a structured literature-oriented analysis rather than an experimental deployment of a specific edge computing framework within a real industrial facility. Therefore, the analysis provides comparative and conceptual findings rather than direct measurements of latency, energy consumption, throughput, or computational performance.

Another limitation is the broad multidisciplinary scope of industrial edge computing. Requirements vary substantially across manufacturing, robotics, energy, transportation, and infrastructure systems. A framework designed for one application may require significant modification when applied to another because latency requirements, device capabilities, communication protocols, and security constraints differ.

The study also focuses on research available through 2024. Consequently, developments published after 2024 are outside the defined literature scope. Furthermore, practical performance can depend strongly on hardware configuration, network topology, workload characteristics, and software implementation. Real-world experimentation is therefore required to validate the conceptual findings presented in this study.

10. Conclusion

Edge computing provides an effective architectural foundation for low-latency industrial decision-making by moving computation closer to industrial devices, machines, sensors, and control systems. Unlike centralized cloud architectures that may require continuous transmission of operational data to remote infrastructure, edge frameworks can process critical information locally and generate rapid decisions. The analysis demonstrates that device edge, gateway edge, distributed edge, AI-enabled edge, and hierarchical edge-cloud architectures can address different industrial requirements depending on latency, computational, scalability, and security considerations.

The integration of artificial intelligence significantly expands the capabilities of industrial edge computing. Edge-based machine learning can support anomaly detection, predictive maintenance, quality inspection, process monitoring, and intelligent control while reducing unnecessary data transmission. However, resource limitations, device heterogeneity, cybersecurity, network reliability, workload management, and scalability remain important implementation challenges. These factors require careful architectural design and continuous system monitoring.

Future industrial edge frameworks should focus on adaptive resource orchestration, lightweight AI, federated learning, secure distributed architectures, digital twins, and autonomous edge-cloud collaboration. Rather than replacing cloud computing, edge computing should operate as part of a hierarchical computational ecosystem in which tasks are allocated according to latency, resource, security, and analytical requirements. With appropriate engineering validation, edge computing can enable industrial systems that are more responsive, resilient, intelligent, efficient, and capable of supporting real-time operational decision-making.

References

1. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646. <https://doi.org/10.1109/JIOT.2016.2579198>
2. Satyanarayanan, M. (2017). The emergence of edge computing. *Computer*, 50(1), 30–39. <https://doi.org/10.1109/MC.2017.9>
3. Chiang, M., & Zhang, T. (2016). Fog and IoT: An overview of research opportunities. *IEEE Internet of Things Journal*, 3(6), 854–864. <https://doi.org/10.1109/JIOT.2016.2584538>
4. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). A survey on mobile edge computing: The communication perspective. *IEEE Communications Surveys & Tutorials*, 19(4), 2322–2358. <https://doi.org/10.1109/COMST.2017.2745201>
5. Mach, P., & Becvar, Z. (2017). Mobile edge computing: A survey on architecture and computation offloading. *IEEE Communications Surveys & Tutorials*, 19(3), 1628–1656. <https://doi.org/10.1109/COMST.2017.2682318>
6. Shi, W., Sun, H., Cao, J., Zhang, Q., & Liu, W. (2016). Edge computing—An emerging computing model for the Internet of Everything era. *Journal of Computer Research and Development*, 53(7), 1440–1451.
7. Abbas, N., Zhang, Y., Taherkordi, A., & Skeie, T. (2018). Mobile edge computing: A survey. *IEEE Internet of Things Journal*, 5(1), 450–465. <https://doi.org/10.1109/JIOT.2017.2750180>
8. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), 1738–1762. <https://doi.org/10.1109/JPROC.2019.2918951>
9. Xu, J., Chen, L., Ren, S., & others. (2018). Online learning for offloading and resource allocation in wireless powered mobile-edge computing systems. *IEEE Transactions on Cognitive Communications and Networking*, 4(3), 435–447.
10. Cao, J., Sun, Y., Zhang, Q., & others. (2020). Edge computing in Industrial Internet of Things: Architecture, advances and challenges. *IEEE Communications Surveys & Tutorials*, 22(4), 2462–2488. <https://doi.org/10.1109/COMST.2020.3009103>
11. Deng, S., Zhang, H., Zhao, Y., Fang, W., Yin, J., Dustdar, S., & Zomaya, A. Y. (2020). Machine learning in mobile edge computing: A survey. *ACM Computing Surveys*, 53(5), Article 102. <https://doi.org/10.1145/3403955>
12. Yu, W., Liang, F., He, X., Hatcher, W. G., Lu, C., Lin, J., & Yang, X. (2018). A survey on the edge computing for the Internet of Things. *IEEE Access*, 6, 6900–6919. <https://doi.org/10.1109/ACCESS.2017.2778504>
13. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. In *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing* (pp. 13–16). ACM. <https://doi.org/10.1145/2342509.2342513>
14. Borsatti, D., Davoli, G., Cerroni, W., & Raffaelli, C. (2021). Enabling Industrial IoT as a service with multi-access edge computing. *IEEE Communications Magazine*, 59(8), 21–27. <https://doi.org/10.1109/MCOM.001.2100006>
15. Ren, J., Yu, G., He, Y., & Li, G. Y. (2019). Collaborative cloud and edge computing for latency minimization. *IEEE Transactions on Vehicular Technology*, 68(5), 5039–5052.

16. Mendez, J., Bierzynski, K., Cuéllar, M. P., & Morales, D. P. (2022). Edge intelligence: Concepts, architectures, applications, and future directions. *ACM Transactions on Embedded Computing Systems*, 21(5), Article 48, 1–41. <https://doi.org/10.1145/3486674>
17. Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., Niyato, D., & Poor, H. V. (2022). Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 23(3), 1622–1658.
18. Kumar, R., & Agrawal, N. (2023). Analysis of multi-dimensional Industrial IoT data in Edge–Fog–Cloud based architectural frameworks: A survey on current state and research challenges. *Journal of Industrial Information Integration*, 32, 100504. <https://doi.org/10.1016/j.jii.2023.100504>
19. Bakhshaei, A., et al. (2022). A survey on task offloading in multi-access edge computing. *Journal of Systems Architecture*, 118, 102225. <https://doi.org/10.1016/j.sysarc.2021.102225>
20. Edge Intelligence for Industrial IoT: Opportunities and limitations. (2024). *Procedia Computer Science*, 232, 397–405. <https://doi.org/10.1016/j.procs.2024.01.039>