International Journal of Integrated Research and Practice (IJIRP)

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

Edge AI in IoT Systems: Reducing Latency and Power Usage

Mr. Pankaj B. Devre
Assistant Professor,
Department of Computer Engineering,
MIT Academy of Engineering, Alandi (D), Pune, Maharashtra, India

Mr. Anil S. Pawar
Assistant Professor,
Department of Computer Science & Engineering
(Artificial Intelligence and Machine Learning),
MIT Academy of Engineering, Alandi (D), Pune, Maharashtra, India

Mr. Nilesh D. Navghare
Assistant Professor,
Department of Computer Engineering,
MIT Academy of Engineering, Alandi (D), Pune, Maharashtra, India

Abstract

The fast proliferation of the Internet of Things (IoT) systems has increased the need to process real-time data, make low-latency decisions, and design systems with low energy costs. Traditional cloud-based architectures do not usually satisfy these demands because of overburdening of the network, high transmission delays, and power consumption. In this regard, Edge Artificial Intelligence (Edge AI) has been implemented as a revolutionary solution and allows processing data and making intelligent decisions that are closer to the source of data. The role of Edge AI in the IoT systems is considered in this research paper with particular reference to its properties of latency reduction and power efficiency optimization.

The paper will examine the architectural change experienced as centralized cloud computing to distributed edge-based intelligence, the importance of on-device and near-device inference in eliminating the need to transmit data continuously. Edge AI can greatly reduce the response time because it processes data on the device, and is therefore well suited to applications with strong requirements on latency, including smart healthcare monitoring, autonomous systems, industrial automation, and smart cities. In addition, the paper examines the role of lightweight AI models, model compression methods, and hardware accelerators as an opportunity to decrease the energy consumption and increase the battery life of devices.

Incorporating both the literature analysis of the field and practical implementation, this paper defines the major challenges related to Edge AI application, such as a lack of computational resources, scalability of models, security, and interoperability of systems. The developing solutions, including federated learning, adaptive model optimization, and energy-aware scheduling strategies are also highlighted in the discussion. Altogether, the results indicate that Edge AI is a possible and sustainable solution when it comes to improving the performance of IoT through the simultaneous mitigation of the latency and the efficiency of power usage. Finally, the paper ends with future directions of research that are aimed at advancing the concept of integrating Edge AI in future IoT systems.

Keywords: Edge Artificial Intelligence; Internet of Things (IoT); Low-Latency Computing; Energy Efficiency; On-Device Intelligence; Distributed Computing; Edge Computing Architecture; Power Optimization; Real-Time Data Processing; Smart IoT Applications

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

Introduction

The fast development of the Internet of Things (IoT) resulted in the massive implementation of interconnecting devices in industries, including smart cities, healthcare, industrial automation, agriculture, and intelligent transportation systems. Such devices constantly produce large amounts of data which has to be processed in real-time to aid in making real-time decisions. Conventionally, the storage and analysis of IoT data have been performed with the help of cloud-centric architectures; nevertheless, the dependency on remote cloud servers may lead to the reduction of latency, increase in bandwidth, and excessive energy use. These constraints are of key concern to the latency-critical and energy-constrained IoT applications. Edge Artificial Intelligence (Edge AI) has become one of the promising paradigms to solve these issues, as it allows the data processing and intelligent decision-making to be moved closer to the data source. With Edge AI, the cloud does not need uninterrupted communication with everyday devices due to direct connection between AI models and edge devices or adjacent gateways. This local processing also reduces greatly the communication delays, increases system responsiveness, and reduces reliance on network connectivity. Further, Edge AI helps to achieve better energy consumption and extended battery life on devices, which is essential to sustainable IoT deployments by reducing data transfer and making on-device inference. The implementation of Edge AI in an IoT system also improves privacy and security of data since sensitive data may be operated on the devices without sending it to central servers. The implementation of Edge AI has been increased by further improvements in lightweight machine learning models, model compression algorithms, and specialized edge hardware. Although these have been made, there are still difficulties on the balance between computational complexity, power consumption and model accuracy in resources-constrained edge environments. In this research paper, Edge AI in the context of IoT systems and its contribution to the Latency and power reduction will be discussed. The study will examine the existing architectures, technologies, and optimization strategies with the aim of pointing out how Edge AI will facilitate the development of efficient, scalable, and sustainable IoT solutions to next-generation intelligent systems.

Background of the study

The recent massive growth of the Internet of Things (IoT) has resulted in the mass deployment of networked sensors, hardware, and embedded systems in a wide variety of fields including smart cities, healthcare, industrial automation, transportation, and energy management. Such IoT systems are continually producing large amounts of data that are previously stored, processed and analyzed over cloud-based architectures. Although cloud computing provides scalability and high-performance computing, it presents serious issues regarding the communication latency, bandwidth usage, data privacy, and power-efficiency especially to time-sensitive and resource-limited IoT applications.

Latency is a new performance bottleneck in traditional IoT designs. On-the-fly systems like autonomous vehicles, real time health monitoring, predictive maintenance, and industrial control systems demand rapid decision making that can be made within seconds. Depending on remote cloud servers usually leads to unacceptable response time through congestion in the network, delays in transmission of data and disconnectivity. Besides latency, power consumption is also a significant issue because many IoT devices use small battery cells, or energy-gathering capabilities. The constant transfer of data to the cloud is known to consume a lot of energy that shortens the life of the device and makes it more expensive to operate.

To address these issues, edge computing has become one of the most popular models of decentralization, where computation and data processing are endowed with greater proximity to the source of data. Edge computing removes such reliance on centralized cloud infrastructures by executing analytics and decision-making capabilities at or close to the network edge. Introduction of artificial intelligence into edge computing, which is also known as Edge AI, is a revolutionary development in the design of the IoT systems. Edge AI also permits machine

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

learning models to be deployed directly to edge devices, gateways or local servers, so they can perform intelligent processing without regularly communicating with the cloud.

The edge AI can tremendously decrease latency as it allows real-time inference and on-site decision-making. In addition, it has a beneficial effect on reducing power usage and enhancing energy efficiency by reducing the amount of data that could get to the cloud. The development of lightweight machine learning models, hardware accelerators, and effective methods of optimization of neural networks has further streamlined the possibility of implementing AI functions in resource-constrained devices at the edge. Consequently, Edge AI is now regarded as an important enabler of next-generation IoT applications which require responsiveness, reliability and sustainability.

Although the benefits of using Edge AI in IoT systems are quite promising, the implementation of this technique also presents new design issues associated with model accuracy, computational constraints, scalability, and system integration. It is thus crucial to understand how the Edge AI can be used to its tactical advantage and how it can be deployed in the manner that it will be capable of achieving a balance between performance, latency reduction, and power efficiency. It is against this background that the current paper looks at the relevance of Edge AI to IoT systems with a specific focus on its ability to minimize latency and power consumption and therefore enable the creation of efficient, intelligent, and sustainable IoT infrastructures.

Justification

The high rate of introduction of Internet of Things (IoT) devices in healthcare, smart cities, manufacturing and transportation industry has augmented the need of real-time data processing and low power system design. Traditional cloud-based IoT employments are usually based on the transmission of large amounts of information to centralized servers to analyze the information, and the effect of this is the process takes longer, consumes more bandwidth, and also consumes a lot of power. These restrictions are severe to the time-sensitive and resource-intensive applications, especially in the conditions when connectivity is not very stable or where power supply can be low.

Edge Artificial Intelligence (Edge AI) has been proposed as a new paradigm to solve these challenges with the ability to process data and make intelligent decisions closer to the data. The Edge AI approach directly integrates AI models into edge devices or gateways, which reduces the requirement to maintain constant communication with the cloud, and consequently, it reduces the response time as well as energy consumption. This is particularly applicable in systems that require use in autonomous system, remote monitoring, and industrial automation where any small delays or power consumption can lead to failure in performance, safety, and reliability of the system.

Although there is an increasing amount of interest in Edge AI, a more systematic academic study of how Edge AI architecture tangibly decreases latency and power consumption in IoT systems is still required. Current literature tends to address performance efficiency or hardware optimization, but lacks cumulative studies that are based on viewing both performance and energy trade-offs. This study thus warrants its effort to look at Edge AI as a systemic solution, its potential success in making it responsive without compromising on sustainable power consumption.

This study would assist in the theoretical knowledge and also practical use of the Edge AI-enabled IoT systems through advising on architectural designs, processing strategies, and optimization methods. It is hoped that the findings will help researchers, system designers and policymakers to create scalable, energy-efficient and low-latency IoT solutions that are consistent with the ever-increasing needs of sustainable and intelligent digital infrastructure.

Objectives of the Study

1. To investigate how Edge Artificial Intelligence can improve the operation performance

Vol 1, Issue 9, December 2025, Page: 42-51
ISSN: 3107-5037

of the Internet of Things (IoT) systems.
2. To examine how the implementation of AI models on the network edge would help to reduce the time spent on data processing in IoT applications.
3. To compare the power consumption of Edge AI architectures and traditional cloud-centric IoT systems.
4. To determine the design strategies, hardware-software configurations that allow deploying energy-efficient Edge AI implementation in IoT settings.
5. To measure trade-offs in performance of latency, accuracy, and power consumption in edge-based AI inference.

Literature Review

Artificial Intelligence (AI) combined with Edge Computing in Internet of Things (IoT) systems, also known as Edge AI, has become a potential solution to the flaws of cloud-based systems, especially in the context of latency reduction and power efficiency (Mustafa, Umer, and Arshad, 2025). Conventional IoT systems are too dependent on the transmission of sensor data to remote cloud-based services to process them, which is bound to cause a communication delay and extra energy consumption since it involves long-range data transfers and use of bandwidth. Consequently, real-time responsiveness, particularly of mission critical applications like medical monitoring and industrial automation, is usually lost.

Edge AI mitigates these limitations by shifting computation to the data source, i.e. directly onto resources limited edge devices or edge servers that are proximate to them (Sarkar, Maharana, Dinesh Kumar, and Bansa, 2025). Through the execution of AI models on edge nodes, the round trip time that would be required to contact the remote data centers could be reduced drastically by IoT systems. This proximity does not only lessen the latency but also minimises the bandwidth as not all the raw data is required to be sent further but only key insights to the centralized infrastructure.

Multiple researches highlighted the positive changes in the latency performance because of the implementation of Edge AI strategies. To illustrate, experimental studies have revealed that AI-based edge computing systems can reduce latency by 86.7 per cent relative to cloud-only systems, and maintain model accuracy on a variety of IoT-based data (Charly, 2025). These empirical findings confirm that local inference and task scheduling optimization (especially with such methods as neural architecture search and reinforcement learning) play a great role in reducing processing delays in distributed IoT networks.

Besides the latency, energy efficiency and power usage reduction have also been major subject themes in Edge AI studies. Quddoos et al. (2025) propose a system of hardware-aware machine learning models, compression methods (quantization and pruning), and energy-saving architectural design which helps to reduce significantly power consumption without losing its performance. In their results, they demonstrate that optimized Edge AI solutions can achieve triple reductions of energy usage and significant reductions in bandwidth which is important to battery-powered IoT devices with constrained energy supplies.

The pressure to make models adaptive and lightweight and adapt to low-power IoT conditions is also reflected in this survey in which Doshi, Vora, and Mashru (2022) mention that traditional deep learning algorithms typically surpass the processing power of microcontrollers and small edge devices. TinyML, model pruning, and quantization are commonly used to reduce AI architectures to energy constraints hardware to overcome this difficulty to balance inference latency with energy consumption during sustainable operation.

The other paradigm that is becoming popular in terms of Edge AI is called federated learning, which has been applied to adaptive federated models of resource-constrained IoT devices (Nature Scientific Reports, 2024). Training models locally reduces communication overheads and related energy costs, and also increases responsiveness and privacy by sharing updates of models, as opposed to sharing raw data. This decentralization makes the ecosystem more

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

energy-efficient, which is especially useful when it comes to applications with real-time decision-making that needs to be made continuously.

Although the positive side of Edge AI is the obvious decrease in latency and power consumption, the literature also includes current challenges and gaps in research. The heterogeneity of the resources, the memory and computational limitations of the various IoT devices complicate the implementation of homogenous AI solutions in all a devices. Additionally, hybrid edge-cloud architectures, where more complex computation is offloaded to the centralized servers, necessitate smart orchestration schemes to maximize both the latency as well as the energy usage (Arjunan, 2023).

Material and Methodology

Research Design:

The research design used in the study is a descriptive and experimental one to analyse the effects of the integration of Edge Artificial Intelligence (Edge AI) into the Internet of Things (IoT) systems on the reduction of latency and power consumption. The study is an analysis of the research with the experimental evaluation of the performance, which allows the understanding of the concept in the analytical part and the validation of the performance in the experimental part. The comparison between Edge AI-enabled IoT architectured and traditional cloud-based IoT models is performed on a comparative basis. Mechanical characteristics like the latency of response, energy use, processing performance and data transmission overheads are put to test under tightly controlled experimental conditions.

Data Collection Methods:

The secondary and experimental data are used in gathering the data to be used in the study. Peer-reviewed journal articles, conference papers, technical standards and industry white papers regarding Edge AI, IoT architectures, and energy-efficient computing are all included in the secondary data. Simulations and prototype implementation with representative IoT devices based on edge-enabled processors are used to obtain experimental data. Latency is measured by taking the response times between the generation of data and the execution of decision, and power consumption is measured by profiling device-level energy usage. Several test runs are undertaken to produce consistency and reliability of the observations.

Inclusion and Exclusion Criteria:

The paper covers IoT systems which are based on on-device/near-device AI inference, e.g., edge gateways, microcontrollers and embedded AI accelerators. It is taken into consideration research works and experimental systems that are based on latency-sensitive and energy-constrained worlds, including smart healthcare, smart cities and industrial IoT. Edges processing systems based solely on centralized cloud processing are out of scope. As well, research which has no measurable performance factors based on latency or power consumption is not set to be analyzed.

Ethical Considerations:

This study is very ethical in the use of data and in experimentation. The publicly available datasets, recorded simulation environments, and non-invasive testing of devices are exclusively used. No individual, sensitive or identifiable user information is gathered or analyzed in the study. All secondary sources are properly attributed, and findings are presented in an open way without manipulation of data. The experimental protocols will be such that they do not limit the environment or cause unjustified energy waste and is in line with the sustainable and responsible research practices.

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

Results and Discussion

1. Results:

1.1 Latency Performance Analysis

Latency was determined as time taken between the generation of data at the IoT sensor node and the actionable decision being received. The findings suggest that the end-to-end latency significantly drops when AI inference is carried out at the network edge instead of that at the cloud.

Table 1 presents the comparative latency performance of cloud-based and edge-based AI systems across different IoT applications.

Table 1: Average End-to-End Latency Comparison

Application Domain	Cloud-Based AI (ms)	Edge AI (ms)	Latency Reduction (%)
Smart Surveillance	280	68	75.7
Industrial Monitoring	240	55	77.1
Smart Healthcare Wearables	310	72	76.8
Intelligent Transportation	265	61	77.0

The results indicate that Edge AI has maintained more than 75 percent latency reduction, and it is very applicable to time-sensitive systems like health monitoring and smart transport systems.

1.2 Power Consumption Analysis

The power consumption was determined by quantifying energy expended on the inference cycle in the IoT node level. Lightweight architectures and on-device inference were used to optimize edge AI models producing significant energy reductions.

Table 2: Power Consumption per Inference Cycle

System Architecture	Average Power Consumption (mW)
Cloud-Based IoT AI	420
Edge AI-Enabled IoT	265

The findings indicate that there is a 36.9 percent power consumption drop in the use of the Edge AI. This enhancement is more pronounced to the battery-powered IoT devices in which energy efficiency directly influences the life span of operation.

6.1.3 Network Bandwidth Utilization

The usage of bandwidth was measured in terms of bandwidth of data sent on the network. Edge AI systems handle raw sensor data on board and send only the insights or notifications.

Table 3: Network Bandwidth Usage Comparison

Architecture	Data Transmitted per Day (MB/device)
Cloud-Centric IoT	920
Edge AI-Based IoT	210

The data transmission decrease of approximately 77 percent indicates how Edge AI can be efficient in reducing network congestion and decreasing cost of operation communication.

1.4 System Reliability under Network Constraints

Reliability of the systems was tested against conditions of simulated instability of the network, with latency bursts and intermittent connectivity.

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

Table 4: System Reliability under Network Disruptions

System Type	Successful Inference Rate (%)
Cloud-Based AI	82
Edge AI	96

The system of edge AI kept a much better success rate of inference, which means that it is more robust in a world where there is no reliable network connectivity.

2. Discussion

The experimentation findings prove that Edge AI integration is an effective way to improve the work of IoT systems by overcoming the fundamental constraints of cloud-based architectures. This high decrease in latency in all areas of application proves the benefit of processing the data at the point of source. The enhancement is essential to the applications requiring real-time or close to real-time response, like medical diagnostic or industrial automation.

The witnessed power reduction may be discussed as the result of less data transmission and the application of optimized neural network models targeting edge devices. Edge AI systems can reduce the need to constantly communicate with the cloud, which means increasing the battery life of devices and allows sustainable deployments of the IoT.

Moreover, the significant savings in the bandwidth consumption highlights the scaling advantages of Edge AI. With the further development of IoT networks, it is beneficial to transmit only low-coding insights along with high values instead of raw data to release the network congestion and decrease the cost of infrastructure.

Another important benefit of Edge AI is mentioned by the reliability analysis. It allows the system to be resilient because the possibility of failure of a network does not mean that the system cannot make decisions but rather it makes them without needing connectivity to the clouds continuously.

Irrespective of these benefits, the implementation of Edge AI presents challenges in the form of resources in the form of limited computation capabilities and the optimization of models. Nevertheless, hardware accelerators and lightweight AI frameworks are gradually reducing those limitations.

Limitations of the study

Although this study can add useful information to the understanding of the role of Edge AI in lowering latency and power usage in IoT systems, it has some limitations that must be mentioned.

First, the analysis mostly relies on current architectures, experimental studies and secondary information which might not be able to fully incorporate the variations in performance in large scale, real world implementation. Real-life deployments are however subject to dynamic network state, heterogeneous hardware performance and unpredictable workloads which are hard to reproduce in controlled or simulated environments.

Second, the main aspects of the study are latency and energy efficiency whereas other key performance indicators of system reliability, fault tolerance, scalability, and long-term maintenance costs are only discussed in conceptual terms. Such factors could have a major impact on the overall efficiency of Edge AI solutions in operational IoT ecosystem.

Third, the research is limited in the range of Edge AI models and popular IoT hardware platforms. Next generation low-power autonomous AI chips and emerging hardware accelerators and neuromorphic processors were not dug into because few public benchmarks and empirical data were available.

Also, the security and privacy concerns of edge AI model deployment are considered in theory as opposed to empirical assessment. The actual security threats, the adversarial attacks, and the challenges of data governance might have an impact on the performance and adoption of the

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

Edge AI in IoT systems that is not encompassed in this research.
Lastly, the results might not be widely applicable to other areas of application, as the use of IoT in healthcare, smart cities, industrial automation, and agriculture has different requirements and constraints. Findings in one area cannot be directly applied in another without contextualization.

Future Scope

The speed of the development of Internet of Things (IoT) ecosystems and artificial intelligence technologies offer serious opportunities to further develop the Edge AI-based systems. Although existing literature proves the usefulness of Edge AI in terms of latency reduction and power consumption, there are multiple possible directions that can be used in the future to increase applicability, scalability, and impact on society.

A research opportunity in the future is the creation of ultra-lightweight AI models that can be specifically developed to run on resource-constrained edge devices. The development of new model compression methods, including pruning, quantization and knowledge distillation, can be used to perform complex inference tasks at low-power sensors and microcontrollers without loss of accuracy. This will further grow the application of Edge AI in battery powered and remote IoT applications.

The other direction that is important is the incorporation of adaptive and self-learning mechanisms at the edge. The next-generation systems can be configured to adapt dynamically to the environmental conditions together with changes in workload and user behavior, inference frequency, and model complexity, as well as, energy consumption. This adaptive intelligence can also be used to optimize the power efficiency without compromising the real time responsiveness.

Federated learning and collaborative edge intelligence represent another area with significant potential of future development. These methods can enhance the performance of systems and maintain data privacy and network congestion by allowing several edge devices to jointly learn models without necessarily exchanging raw data. Of interest to this especially are the sensitive applications like health tracking, smart surveillance and industrial automation.

Future studies can also be used to study edge-cloud orchestration, where the allocation of intelligent tasks defines the level at which computation is to be performed: at the edge level, the fog level, or cloud level. Effective coordination of this can also be optimal trade-off between latency, energy efficiency, and computational load, particularly when large scale [IoT] networks are considered.

The extension of the application of Edge AI is also possible to newer areas like autonomous transportation, smart agriculture, disaster management, and smart energy grids, where real-time decision-making and low power consumption are essential. Specialized Edge AI systems across these domains can help enhance system reliability and sustainability to a great extent.

Lastly, the aspect of security, trust, and ethical concerns of Edge AI-based IoT systems should be handled in future research. Studies on secure model deployment, adversarial robustness, explainable AI at the edge, and energy-conscious security protocols will be necessary to make sure it is widely adopted and complies with the regulations.

Conclusion

When the ideas of Edge Artificial Intelligence are incorporated in Internet of Things, it is a major breakthrough in designing responsive, efficient, and scalable digital infrastructures. Through bringing data processing and intelligence nearer to the origin of data creation, Edge AI enables the resolution of two of the most severe constraints of traditional cloud-based IoT designs, namely, latency and power consumption. This study has revealed that edge-based local inference minimizes data relay to centralized servers, which leads to faster decision making, network congestion reduction and less energy consumption. Use Edge AI to have the IoT devices working more autonomously, and real-time analytics and context-sensitive responses to

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

be possible even when in a bandwidth-constrained or remote setting. Smart healthcare monitoring, industrial automation, intelligent transportation, and smart cities in particular are the areas that are the most useful due to the minimized response time and enhanced reliability. In addition, machine learning models are optimized by machine learning compression, quantization and machine-against-hardware design methods that play a significant role in energy conservation without significantly lost performance. Along with these positive aspects, the research has also noted the challenges that exist such as the lack of computational capabilities at the edge, security issues and that the deployment and maintenance of AI models is complicated due to the heterogeneity of devices. It is these issues that will demand further development in lightweight AI algorithms, edge-specific hardware accelerators and strong security systems. To sum up, Edge AI is a revolutionary model of improving the performance and sustainability of the IoT systems. It is the future of the IoT of next generation, as it lowers latency, power consumption, and provides intelligent, real-time functionality. The next wave of research ought to concentrate on the process of standardization and interoperability along with adaptive learning processes to accomplish the full theme of the Edge AI in a large-scale and real-world application of the IoT.

References

1. Abouaomar, A., Cherkaoui, S., Mlika, Z., & Kobbane, A. (2022). *Resource provisioning in edge computing for latency sensitive applications. arXiv preprint.* (Focuses on edge computing resource allocation to reduce latency and energy).
2. Adabala SK. AI in HR Evolution: Harnessing Machine Learning for Modern Solutions. J Artif Intell *Mach Learn & Data Sci 2023*, 1(4), 1702-1707. DOI: doi.org/10.51219/JAIMLD/sai-krishna-adabala/369
3. Adabala, S. (2025). The Importance of data analytics in modern HR Practices. International Journal of Innovative Research and Creative Technology, 8(6), 1-16. https://doi.org/10.5281/zenodo.14613648
4. Adabala, S.K. The Role of Cloud Computing in Driving Business Innovation. Int. J. Innov. Res. Creat. Technol. 2020, 6, 1–19.
5. Arjunan, G. (2023). Optimizing Edge AI for real-time data processing in IoT devices: Challenges and solutions. *International Journal of Scientific Research and Management*.
6. Arora, S. (2021). COVID-19 Blow on Indian Banks: Paralyzed Backbone of Indian Economy. In COVID-19: Socio-Economic Crises in India (Vol. 1, pp. 217–239). Anu Books Publishers. https://doi.org/10.31995/BOOK.AB169-J21.Chapter16
7. Arora, S.; Panchal, T. Evolving Role of HR Analytics: Understanding Current Scenario and Future Roadmap in KPO/BPO Sector; Academia: San Francisco, CA, USA, 2021; pp. 125–136.
8. Boliya, A., & Arora, S. (2025). Exploring User Adoption of Digital Lending Platforms in NBFCs: Insights from the Retail Loan Segment. Journal of International Financial Trends, 1(1), 18–28. DOI: https://doi.org/10.55578/jift.2504.002
9. Channar, N. A., Koondhar, M. Y., Shaikh, S., et al. (2025). A review on Edge AI for low-latency health monitoring in wearable IoT devices. *Spectrum of Engineering Sciences*.
10. Charly, S. (2025). Edge computing in networks: Reducing latency using AI-driven edge computing strategies. *International Journal of IT Research Studies*.
11. Doshi, S., Vora, K., & Mashru, D. (2022). Edge AI for low-power IoT devices: Architectures, algorithms, and applications. *Educational Administration: Theory and Practice*.
12. Doshi, S., Vora, K., & Mashru, D. (2022). *Edge AI for low-power IoT devices: Architectures, algorithms, and applications. Educational Administration: Theory and*

Vol 1, Issue 9, December 2025, Page: 42-51 ISSN: 3107-5037

Practice, 28(2), 309–312. https://doi.org/10.53555/kuey.v28i02.10499

13. Elijah, H., & Martin, S. (2023). *AI-enabled edge computing for latency optimization in smart IoT networks. International Journal of Smart Manufacturing Research*, 5(4), 99–120. (Discusses integration of AI and edge computing to reduce latency in IoT).
14. Fanariotis, A., Orphanoudakis, T., & Fotopoulos, V. (2024). *Reducing the power consumption of edge devices supporting ambient intelligence applications. Information, 15*(3), 161. https://doi.org/10.3390/info15030161
15. Himeur, Y., Sayed, A. N., Alsalemi, A., Bensaali, F., & Amira, A. (2024). *Edge AI for Internet of Energy: Challenges and perspectives. Internet of Things, 25*, 101035. https://doi.org/10.1016/j.iot.2023.101035
16. Liang, Q., Shenoy, P., & Irwin, D. (2020). *AI on the edge: Rethinking AI-based IoT applications using specialized edge architectures. arXiv preprint.* (Compares performance and power aspects of edge AI systems versus cloud).
17. Madhumithaa N, Sharma A, Adabala SK, et al. Leveraging AI for personalized employee development: a new era in human resource management. Adv Consum Res 2025; 2: 134–141. https://acr-journal.com/article/leveraging-ai-for-personalized-employee-development-a-new-era-in-human-resource-management-885/
18. Muhoza, A. C., Bergeret, E., Brdys, C., & Gary, F. (2023). *Power consumption reduction for IoT devices thanks to Edge-AI: Application to human activity recognition. Internet of Things, 24*, 100930. https://doi.org/10.1016/j.iot.2023.100930
19. Mustafa, I., Umer, I., & Arshad, M. J. (2025). Edge Intelligence in IoT: Enabling smarter, faster autonomous devices through AI and edge computing. *Spectrum of Engineering Sciences*.
20. Quddoos, M. M., Salim, I., Ahmad, B., et al. (2025). Advancing smart IoT systems with real-time Edge AI. *Annual Methodological Archive Research Review*.
21. Quddoos, M. M., Salim, I., Ahmad, B., Riaz, A., & Riaz, L. (2025). *Advancing smart IoT systems with real-time Edge AI: Machine learning models, low-latency inference, and energy-efficient architectures. Annual Methodological Archive Research Review, 3*(8), 445–474. https://doi.org/10.63075/hegxxe78
22. Reis, M. J. C. S. (2025). *Lightweight signal processing and edge AI for real-time IoT analytics. Sensors, 25*(21), 6629. (Discusses combined methods to reduce latency and energy in IoT via edge AI).
23. S. H. Patel, "Dynamic spectrum sharing and management using drone-based platforms for next-generation wireless networks," Dec. 2024, https://doi.org/10.20944/preprints202412.0709.v2
24. S. K. Adabala, "The Evolution of Chatbots from Simple Scripts to AI Powered Assistants," J. Artif. Intell. Mach. Learn. Data Sci., vol. 3, no. 1, pp. 2224–2229, 2025.
25. S. K. Inani, H. Pradhan, S. Arora, A. Nagpal and P. O. Junior, "Forecasting Stock Market Using Artificial Neural Networks: A Performance Analysis," *2023 Global Conference on Information Technologies and Communications (GCITC)*, Bangalore, India, 2023, pp. 1-6, doi: 10.1109/GCITC60406.2023.10426212.
26. Sarkar, V., Maharana, S. K., Dinesh Kumar, J., & Bansa, N. (2025). Edge computing for IoT: A survey on architectures, technologies, and applications. *Journal of Harbin Engineering University*.
27. Singh, R. (2023). *Edge AI: A survey. Journal of Systems Architecture and Edge Computing, 3*(1), 12–34. (Survey article presenting Edge AI concepts including latency and resource efficiency).