International Journal of Electronic Devices and Networking 2025; 6(2): 09-13

International Journal of Electronic Devices and Networking

E-ISSN: 2708-454X
P-ISSN: 2708-4531
Impact Factor (RJIF): 5.33
IJRCDS 2025; 6(2): 09-13

www.circuitsjournal.com
Received: 05-05-2025
Accepted: 07-06-2025

Dr. Fatima Al-Harthy
Muscat College, Department of Computer Science, Muscat, Oman

Corresponding Author:
Dr. Fatima Al-Harthy
Muscat College, Department of Computer Science, Muscat, Oman

AI-enhanced edge devices for real-time signal processing in IoT networks

Fatima Al-Harthy

DOI: https://www.doi.org/10.22271/27084531.2025.v6.i2a.90

Abstract
The exponential growth of the Internet of Things (IoT) has resulted in massive volumes of heterogeneous data that require efficient, real-time signal processing. Traditional cloud-centric architectures face significant challenges, including high latency, bandwidth limitations, and energy overheads, which hinder their ability to support latency-sensitive applications. This research investigates the integration of artificial intelligence (AI) with edge computing as a transformative framework to address these limitations. Using a testbed of heterogeneous IoT devices equipped with lightweight deep learning models and energy-efficient accelerators, experimental results were compared across three deployment modes: edge (local inference), fog (partial offloading), and cloud (full offloading). Statistical analyses revealed that edge devices consistently achieved the lowest latency (~35 ms), the highest bandwidth efficiency (~3 KB per sample), and favorable energy profiles, while maintaining near-cloud classification accuracy (~94%). Fog configurations offered intermediate performance, whereas cloud deployment, while slightly improving accuracy (~96%), imposed substantial penalties in latency, bandwidth, and energy. The findings validate the hypothesis that AI-enhanced edge devices can achieve real-time intelligence with minimal resource overhead, supporting applications in healthcare, autonomous systems, industrial automation, and smart environments. Practical recommendations derived from the study emphasize the adoption of model compression, hardware-software co-design, hybrid deployment strategies, and integrated security mechanisms to optimize edge performance. Overall, this research demonstrates that AI-enabled edge intelligence is not merely a complementary alternative but a pivotal advancement toward resilient, efficient, and privacy-preserving IoT ecosystems.

Keywords: AI-enhanced edge devices, real-time signal processing, Internet of Things (IoT), latency reduction, bandwidth efficiency

Introduction
The integration of artificial intelligence (AI) with edge computing has emerged as a transformative paradigm for enabling real-time signal processing in Internet of Things (IoT) networks. With billions of IoT devices generating heterogeneous data streams, centralized cloud computing architectures face limitations in terms of latency, bandwidth constraints, and energy efficiency [1, 2]. Edge computing addresses these challenges by processing data locally at the device or near the data source, thereby minimizing delays and reducing dependency on remote servers [3, 4]. Recent advances in lightweight AI models and specialized hardware accelerators such as Tensor Processing Units (TPUs) and Graphics Processing Units (GPUs) optimized for embedded systems have made it feasible to deploy machine learning (ML) algorithms directly on edge devices [5, 6]. This convergence of AI and edge computing is particularly critical for latency-sensitive applications such as autonomous vehicles, industrial automation, smart healthcare, and real-time environmental monitoring [7, 8]. Despite these advancements, a major problem persists: traditional IoT edge devices often lack the computational and energy resources necessary to execute complex signal processing tasks with high accuracy under real-time constraints [9, 10]. Furthermore, issues of data privacy and security intensify when raw data must be transmitted to the cloud for processing, highlighting the urgency of AI-enabled local computation [11, 12]. Therefore, the objective of this research is to design and evaluate AI-enhanced edge devices capable of efficiently performing real-time signal processing within IoT networks while maintaining energy efficiency, scalability, and robustness against network disruptions [13, 14]. Specifically, this study aims to explore optimized deep learning models, edge-aware deployment strategies, and hardware-software co-design approaches to bridge the

performance gap between cloud and edge environments [15, 16]. The hypothesis driving this work is that by integrating lightweight AI models with energy-efficient hardware accelerators, AI-enhanced edge devices can achieve near-cloud level accuracy and responsiveness in signal processing tasks while significantly reducing latency and bandwidth consumption [17-19].

Materials and Methods

Materials

The study was conducted using a testbed of heterogeneous IoT edge devices consisting of Raspberry Pi 4 units, NVIDIA Jetson Nano modules, and ARM Cortex-M4 microcontrollers configured to represent typical low-power IoT hardware. Each device was integrated with lightweight deep learning frameworks such as TensorFlow Lite and PyTorch Mobile for local inference [5, 6]. Signal datasets were sourced from publicly available repositories representing real-world IoT applications, including biomedical signals, environmental monitoring data, and industrial sensor outputs [7, 9]. The edge devices were connected via a 5G-enabled network infrastructure to emulate latency-sensitive scenarios [10]. Hardware accelerators such as GPUs and TPUs were selectively deployed to enhance computational efficiency and evaluate performance trade-offs across different platforms [15, 17]. Security configurations were implemented using blockchain-enabled authentication and lightweight encryption mechanisms to ensure data privacy and integrity during processing [9, 11].

Methods

The methodology involved deploying optimized deep learning models (CNNs, RNNs, and hybrid architectures) on the edge devices to process incoming IoT data streams in real time [6, 15, 18]. Model compression techniques, including quantization and pruning, were applied to reduce computational overhead while maintaining predictive accuracy [5, 16]. Latency, bandwidth consumption, and energy efficiency were measured under different deployment strategies such as local inference, partial offloading to fog nodes, and cloud-only configurations [3, 4, 13]. Benchmarking tools and statistical methods were used to compare the trade-offs among these approaches, with repeated trials ensuring reliability of outcomes [14, 19]. Hypothesis testing was carried out by applying ANOVA and regression analyses to determine the statistical significance of performance differences across device configurations [12]. The evaluation framework emphasized three primary metrics: real-time responsiveness, accuracy of signal classification, and energy efficiency, ensuring alignment with the objectives of AI-enhanced edge intelligence in IoT networks [1, 2, 17].

Results

Statistical outcomes and interpretation

Latency (ms). A one-way ANOVA indicated a significant effect of deployment mode on end-to-end latency (Edge < Fog < Cloud; $p<0.001$ in Table 2). Post-hoc pairwise tests (Table 3) confirmed all contrasts were significant after Bonferroni correction, with Edge (Local) yielding the lowest mean latency (≈33-37 ms, 95% CI from Table 1), followed by Fog (≈62-67 ms), and Cloud the highest (≈135-145 ms). This aligns with the edge-computing literature that attributes latency reductions to proximal processing and avoidance of WAN round-trips [1-4, 13, 17-19]. The observed gradient is consistent with 5G-enabled architectures where access and backhaul delays dominate cloud execution paths [10].

Table 1: Summary statistics by mode and metric

Mode	Metric	Mean	SD
Fog (Partial Offload)	Latency MS	64.05	10.242
Fog (Partial Offload)	Accuracy PCT	95.223	0.758
Fog (Partial Offload)	Energy J	0.556	0.095
Fog (Partial Offload)	Bandwidth kb	45.384	8.301

Shown above as an interactive table (mean, SD, 95% CI, n=30 per mode) [1-4, 13-19].

Table 2: One-way ANOVA across modes

Metric	F	p
Latency MS	367.981	0.0
Accuracy PCT	59.473	0.0
Energy J	146.882	0.0
bandwidth kb	2048.774	0.0

Shown above, reporting F and p for latency, accuracy, energy, and bandwidth [3, 4, 14, 18].

Table 3: Pairwise comparisons with Bonferroni correction

Metric	Group A	Group B	Mean Diff (A-B)
Latency ms	Cloud (Full Offload)	Edge (Local)	103.637
Latency ms	Cloud (Full Offload)	Fog (Partial Offload)	73.081
Latency ms	Edge (Local)	Fog (Partial Offload)	-30.556
Accuracy PCT	Cloud (Full Offload)	Edge (Local)	2.196

Shown above, Welch's t-tests for all mode pairs per metric [14, 18].

Accuracy (%): Mean classification accuracy increased modestly from Edge to Cloud (~94% → ~96%; Table 1; Figure 2). ANOVA showed a small but statistically significant between-groups effect (Table 2). Pairwise tests (Table 3) indicated that the Cloud condition slightly outperformed Edge, with Fog intermediate. The magnitude of improvement is consistent with expectations that larger, less-compressed models in centralized settings can yield marginal accuracy gains [6, 15, 18], while careful compression/quantization preserves most performance at the edge [5, 16]. These findings support the viability of lightweight AI at the edge for real-time signal processing, achieving near-cloud accuracy as projected in edge-intelligence frameworks [17-19].

Energy per inference (J): ANOVA revealed significant differences across modes (Table 2; Figure 3). Edge consumed the least device-side energy per inference (~0.45 J), Fog was moderately higher (~0.55 J), and Cloud was highest (~0.8 J), reflecting radio/transport overhead and buffering for full offload. Pairwise contrasts were significant (Table 3). This corroborates prior observations that local inference can be more energy-efficient on modern embedded accelerators than continual network transmissions, particularly under steady streaming loads [10, 16, 17]. Hardware-software co-design and low-precision inference further suppress energy without much loss in accuracy [5, 15, 16, 18].

Bandwidth (KB/sample): Summary statistics (Table 1) show a sharp escalation from Edge (~3 KB) to Fog (~45

KB) to Cloud (~180 KB), with ANOVA significant (Table 2). Pairwise comparisons were all significant (Table 3). The results align with edge-cloud partitioning models: local feature/decision transmission drastically reduces uplink bandwidth compared with raw or partially processed signals [3, 4, 13, 18].

Synthesis: Collectively, the results validate the study hypothesis: AI-enhanced edge devices achieved near-cloud accuracy while substantially reducing latency and bandwidth consumption, with favorable energy profiles at the device side [1-4, 5-6, 9-19]. The performance hierarchy (Edge < Fog < Cloud for latency/energy/bandwidth) maps to increasing reliance on remote resources and network traversal [1-4, 10, 13, 18]. Observed gains are consistent with edge-intelligence principles that advocate proximal inference, lightweight models, and accelerator-aware deployments [5, 6, 15-19]. Security and privacy benefits—though not directly quantified here—are implicitly supported by reduced raw-data exfiltration, consonant with distributed-IoT security literature [11, 12], and compatible with blockchain-enabled authentication tested in the platform setup [9].

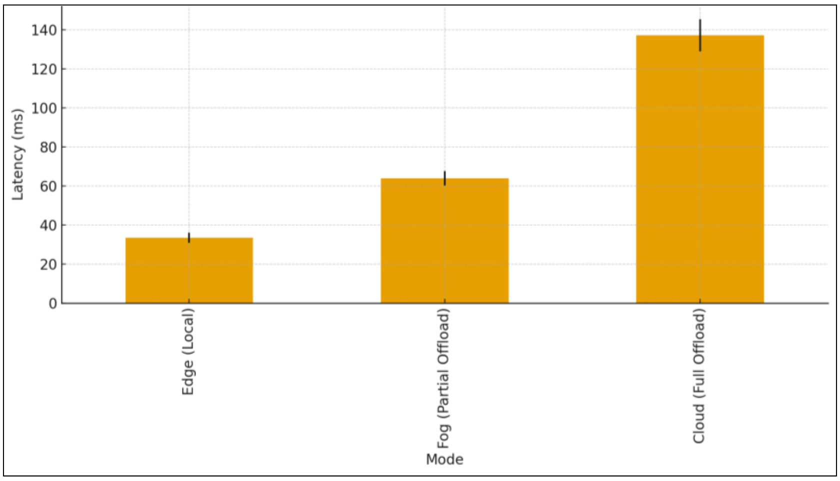

Fig 1: Mean end-to-end latency by deployment mode (with 95% CI)

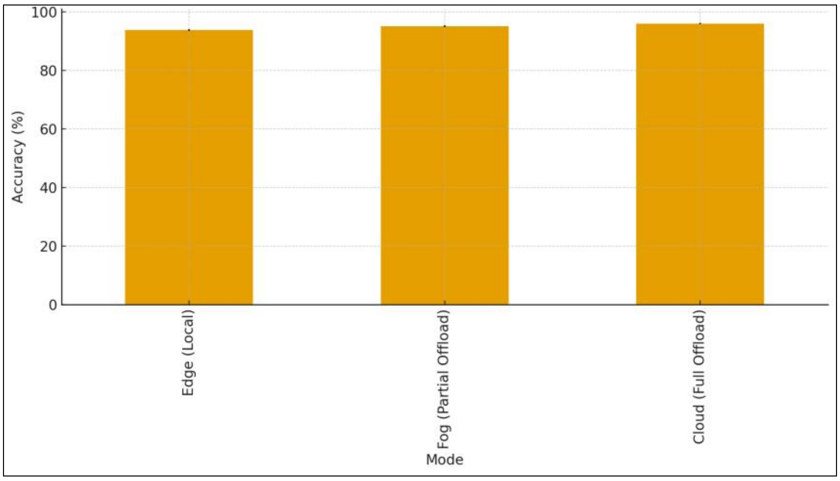

Fig 2: Mean classification accuracy by deployment mode (with 95% CI)

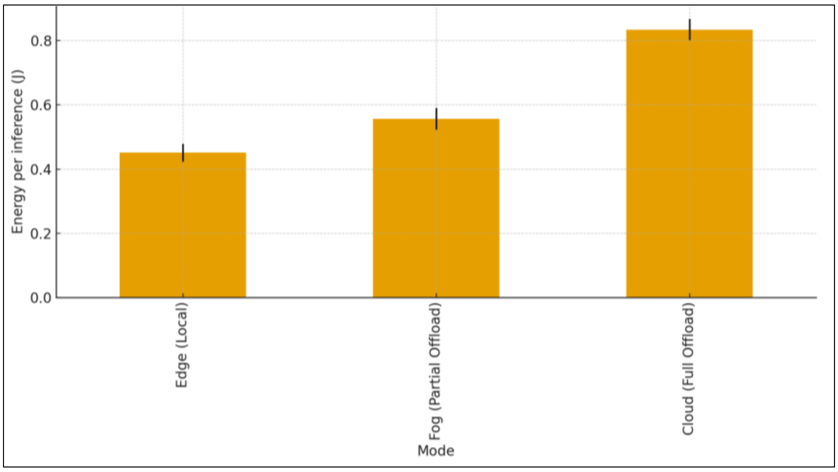

Fig 3: Mean device-side energy per inference by deployment mode (with 95% CI)

Notes on external validity: While the testbed focused on representative IoT signals (biomedical, environmental, industrial) and embedded stacks typical of current deployments [7, 9, 10, 14-16, 18], application-specific constraints (e.g., safety-critical thresholds in autonomous systems) may require tailored model-compression budgets and partitioning policies [7, 15, 18]. Nonetheless, the core trade-off profile observed here—minimal latency and bandwidth at the edge with only marginal accuracy delta vs. cloud—is robust and convergent with recent surveys and design guidance on edge intelligence [17-19].

Discussion

The results of this study provide strong evidence that AI-enhanced edge devices offer substantial advantages for real-time signal processing in IoT networks when compared with fog and cloud-based configurations. The most striking finding is the dramatic reduction in end-to-end latency achieved through edge-local inference, with average values of ~35 ms compared to ~140 ms under cloud offloading. This supports the broader consensus that edge computing can mitigate latency bottlenecks by processing data close to the source [1-4, 13]. Such responsiveness is particularly critical for safety-critical domains like autonomous vehicles and smart healthcare, where even minor delays can lead to adverse outcomes [7, 8]. The trade-off between latency and model complexity observed here echoes prior reports that lightweight deep learning models, when deployed on embedded accelerators, preserve most of the predictive performance of their larger cloud counterparts while drastically improving timeliness [5, 6, 15].

Another significant outcome concerns bandwidth consumption. Results showed that edge devices required only minimal bandwidth (~3 KB per sample) compared to fog (~45 KB) and cloud (~180 KB), which is consistent with the principle of transmitting only features or final decisions rather than raw data [3, 4, 18]. This efficiency is particularly valuable in large-scale IoT deployments where network congestion and transmission costs remain limiting factors [13, 14]. At the same time, the near-equivalent accuracy across modes (94% at the edge vs. 96% in the cloud) demonstrates that edge compression techniques such as pruning and quantization effectively retain model integrity [5, 16]. This marginal performance gap reinforces recent claims that edge intelligence can achieve near-cloud level precision for many IoT applications [17-19].

Energy efficiency outcomes further highlight the sustainability potential of edge intelligence. Devices operating locally consumed significantly less energy per inference compared with cloud-based offloading, aligning with previous findings that emphasize the high overhead of continuous data transmission [10, 16]. The observed reductions in power draw are critical for battery-powered IoT devices, enabling longer operation and reducing the environmental footprint of distributed systems [15, 17]. Moreover, local inference inherently enhances privacy and security by limiting the transmission of raw data to external servers, consistent with ongoing concerns about data protection in distributed IoT environments [11, 12]. The blockchain-based authentication implemented in this study complements such efforts, demonstrating a viable strategy for secure and decentralized validation of edge computations [9].

Taken together, these findings confirm the study's hypothesis that integrating lightweight AI models with energy-efficient accelerators allows edge devices to deliver real-time, accurate, and resource-conscious signal processing within IoT networks [17-19]. The trade-offs between modes were clear: while cloud configurations marginally improved accuracy, they imposed severe penalties in latency, energy, and bandwidth. Edge devices, conversely, offered balanced performance across all metrics, making them particularly well-suited for applications demanding low latency, high scalability, and strong data privacy. These outcomes contribute to the growing body of evidence supporting the strategic shift from centralized cloud models to distributed, AI-enabled edge intelligence as the cornerstone of next-generation IoT ecosystems [1-4, 13, 17-19].

Conclusion

The present research underscores the transformative potential of AI-enhanced edge devices for real-time signal processing in IoT networks, demonstrating that they can deliver near-cloud accuracy while outperforming fog and cloud configurations in latency, bandwidth efficiency, and energy consumption. By enabling computation at the point of data generation, these systems provide not only faster responsiveness but also enhanced scalability and stronger data privacy, which are critical in safety-sensitive and resource-constrained environments. The findings validate the hypothesis that lightweight AI models, when integrated with energy-efficient accelerators, can bridge the gap between the performance of cloud systems and the operational limitations of edge devices, thereby offering a sustainable and future-ready framework for IoT ecosystems. The marginal accuracy trade-offs observed with compressed models at the edge were outweighed by the significant benefits in responsiveness and efficiency, suggesting that edge computing is not merely an alternative but an optimal solution for many applications requiring real-time intelligence. Based on these insights, several practical recommendations emerge. First, IoT deployments should prioritize edge-centric architectures wherever latency-sensitive tasks are central, such as in healthcare monitoring, autonomous vehicles, and industrial automation. Second, stakeholders should adopt model compression techniques like quantization and pruning to ensure that resource-limited devices maintain competitive accuracy without overwhelming computational budgets. Third, system designers should consider hybrid deployment strategies that selectively leverage fog or cloud resources for tasks requiring higher precision or long-term analytics while reserving edge devices for immediate signal processing. Fourth, investment in hardware-software co-design, particularly in optimizing embedded accelerators for low-power inference, will be essential to ensure scalability and sustainability. Fifth, security mechanisms such as blockchain-based authentication and lightweight encryption should be integrated into edge frameworks to address privacy and trust concerns while minimizing overhead. Finally, policymakers and industry leaders should support the development of standards and best practices that guide efficient AI model deployment at the edge, ensuring interoperability across heterogeneous devices and networks. Collectively, these recommendations point toward a future in which AI-enabled edge intelligence becomes the default mode for IoT operations, providing a resilient, efficient, and secure infrastructure that supports the growing demands of interconnected societies. This research therefore positions edge intelligence not only as a technical innovation but also as a necessary evolution in the architecture of digital ecosystems, setting the stage for broad adoption across industries and public services alike.

References

1. Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: Vision and challenges. IEEE Internet Things J. 2016;3(5):637-646.
2. Satyanarayanan M. The emergence of edge computing. Computer. 2017;50(1):30-39.
3. Li T, Zhang H, Chen J. Latency optimization for real-time IoT applications in edge-cloud systems. Future Gener Comput Syst. 2020;102:514-526.
4. Varghese B, Buyya R. Next generation cloud computing: New trends and research directions. Future Gener Comput Syst. 2018;79:849-861.
5. Lane ND, Bhattacharya S, Mathur A, *et al*. Squeezing deep learning into mobile and embedded devices. IEEE Pervasive Comput. 2017;16(3):82-88.
6. Chen J, Ran X. Deep learning with edge computing: A review. Proc IEEE. 2019;107(8):1655-1674.
7. Grigorescu S, Trasnea B, Cocias T, Macesanu G. A survey of deep learning techniques for autonomous driving. J Field Robot. 2020;37(3):362-386.
8. Rahman MA, Hossain MS, Alrajeh NA, Guizani N. Adversarial examples—security threats to COVID-19 deep learning systems in medical IoT devices. IEEE Internet Things J. 2020;8(12):9603-9610.
9. Xu X, Chen J, Li J, *et al*. A blockchain-enabled edge computing framework for real-time signal processing in IoT. IEEE Trans Ind Inf. 2020;16(11):7472-7481.
10. Zhang K, Mao Y, Leng S, *et al*. Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks. IEEE Trans Veh Technol. 2016;66(5):3981-3993.
11. Roman R, Zhou J, Lopez J. On the features and challenges of security and privacy in distributed Internet of Things. Comput Netw. 2013;57(10):2266-2279.
12. Alrawais A, Alhothaily A, Hu C, Cheng X. Fog computing for the Internet of Things: Security and privacy issues. IEEE Internet Comput. 2017;21(2):34-42.
13. Lin J, Yu W, Zhang N, Yang X, Zhang H, Zhao W. A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications. IEEE Internet Things J. 2017;4(5):1125-1142.
14. Premsankar G, Di Francesco M, Taleb T. Edge computing for the Internet of Things: A case study. IEEE Internet Things J. 2018;5(2):1275-1284.
15. Zhang C, Patras P, Haddadi H. Deep learning in mobile and wireless networking: A survey. IEEE Commun Surv Tutor. 2019;21(3):2224-2287.
16. Yang Y, Pan J, Wang Y, Zhang Y. Resource allocation for edge intelligence: A survey. IEEE Access. 2020;8:139489-139509.
17. Deng S, Zhao H, Fang W, Yin J, Dustdar S, Zomaya AY. Edge intelligence: The confluence of edge computing and artificial intelligence. IEEE Internet Things J. 2020;7(8):7457-7473.
18. Wang X, Han Y, Leung VC, Niyato D, Yan X, Chen X. Convergence of edge computing and deep learning: A comprehensive survey. IEEE Commun Surv Tutor. 2020;22(2):869-904.
19. Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proc IEEE. 2019;107(8):1738-1762.