Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 5, 2025

EDGE INTELLIGENCE IN IOT: ENABLING SMARTER, FASTER AUTONOMOUS DEVICES THROUGH ARTIFICIAL INTELLIGENCE AND EDGE COMPUTING

Idrees Mustafa[*1], Imran Umer[2], M. Junaid Arshad[3]

[*1]*Dept. of Computer Science University of Engineering and Technology Lahore, Pakistan*
[2]*Dept. of Data Science University of Engineering and Technology Lahore, Pakistan*
[3]*Dept. of Computer Science University of Engineering and Technology Lahore, Pakistan*

[*1]2024mscs12@student.uet.edu.pk, [2]2023msds02@student.uet.edu.pk, [3]junaidarshad@uet.edu.pk

DOI: https://doi.org/10.5281/zenodo.15411631

Keywords
Edge Intelligence (EI), Edge Computing, Federated Learning, Latency Reduction, Energy Efficiency, Privacy Preservation, 6G Networks.

Article History
Received on 05 April 2025
Accepted on 05 May 2025
Published on 14 May 2025

Corresponding Author: *
Idrees Mustafa

Abstract
The intersection of Edge Computing and Artificial Intelligence represented by Edge Intelligence (EI) is paving a new road for the Internet of Things (IoT), eliminating its main shortcoming, i.e. the latency, the bandwidth constraints, and the privacy concerns of the traditional cloud systems. In this paper, we have discussed principles, challenges and applications of EI, focusing on its overall contribution to real-time secure and energy-efficient operations in healthcare, smart cities, industrial IOT and autonomous vehicles. Federated learning, lightweight AI models, and hybrid edge-cloud architectures are analyzed on the grounds of their key technological advancement on how energy and scalability restrictions can be overcome. Finally, integration of 6G networks with blockchain technology along with an ethical AI framework is proposed as a path to enable future capabilities. This work intends to point researchers, developers, and policymakers in the direction of adopting security and sustainability in EI system by providing a comprehensive survey of the available solutions and future trends.

INTRODUCTION

The Internet of Things has revolutionized technological usage through smart networking by creating smart city control methods, individual healthcare solutions, and industrial robotic functionalities. Traditionally, most IoT data management pro- cesses are linked to centralized cloud-based platforms for operations. Due to its large data transmission requirements and centralized data storage, this solution presents privacy issues because its usual latency period ranges between 100 to 500 milliseconds while consuming significant bandwidth [1]. However, the implementation of telemedicine services and industrial safety systems requires low latency performance, which must provide results in a few milliseconds of delay.

The approach of edge computing shifts data processing operations from IoT sensors, cameras and wearable technology into a position near their source of data. This system was con- ceived with the purpose of solving major issues. Technological processes in edge computing systems move data processing functions to locations that are nearer to users for improved latency performance. It has reduced the need for distant cloud servers, improving data privacy levels while helping to reduce network traffic [1,2].

The integration of Artificial Intelligence (AI) and Machine Learning (ML) in IoT systems functions through edge computing environments for their principal operational purpose. The main objective of our paper involves describing Edge Intelligence on a conceptual basis. Edge intelligence handles real-time analysis of local data through the operation of light-weight AI models that function directly on constrained edge devices. Removing delays from cloud systems makes the system operational with better performance rates and enhanced security capabilities [2,3]. The implementation of Edge Intelligence in Barcelona's smart cities has led to several positive outcomes exemplified by fast and beneficial traffic system management[3]. Furthermore, the ECG diagnostic devices run analysis applications at 98% accuracy through real-time operations without compromising users' privacy[4].

Edge Intelligence implementation continues to meet multiple hurdles which prevent deployment while showing current advancements. The deployment of Edge Intelligence systems encounters the biggest barriers through power inefficiencies, security challenges and AI model discrimination problems. Improved privacy in Federated learning systems causes per- formance issues due to prediction model distortions that occur from unbalanced data across edge devices [4]. 6G network communication developments are vital to Edge Intelligence advancements since these networks will enable 1Tbps speeds along with 0.1 millisecond latency improvements.

This review paper discusses Edge Intelligence not only at theoretical level but also at technological level and application level. It tries to integrate lightweight artificial intelligence models with edge computing for improved IoT system functionality. In this work, the key challenges such as energy efficiency, latency-accuracy tradeoffs, security, and scalability are studied systematically and some solutions by employing federated learning, hybrid architectures and advanced cryptographic methods are proposed. In addition, this work discovers important answers to the future of developments, such as the adoption of 6G networks, blockchain technology and ethical framework of AI. This paper attempts to act as a resource for future researchers, practitioners and policymakers who would like to produce secure, efficient and scalable EI systems in next generation IoT environments by integrating recent research findings with industry.

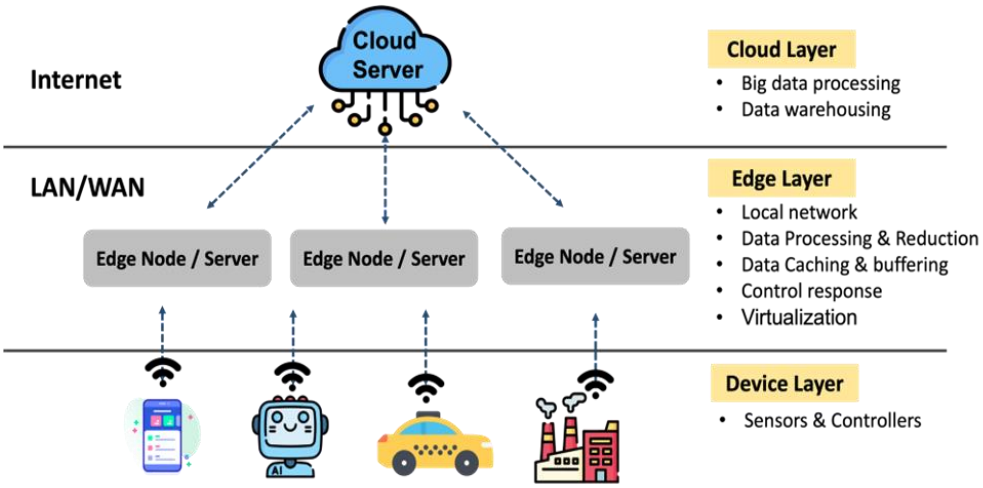

Fig. 1. Architecture of Edge Computing[3]

I. Foundations of Edge Intelligence

A strong technological base is essential for the combined operation of edge computing along with artificial intelligence. The system begins operation at edge devices before moving to edge servers which finally leads towards cloud architecture as described in figure 1. Edge devices have two primary tasks: they carry out limited data computation using sensors, smart cameras, and wearable technology. AWS IoT Greengrass and other intermediate-level edge servers

function with routers as they efficiently handle swift data processing from multiple edge devices. Cloud-based technology achieves improved data speed transfers and lessened bandwidth usage when enhanced with this innovation. The placement of edge servers demonstrates a reduction in cloud bandwidth usage at a rate of 40% based on recent research findings[6].

The deployment of artificial intelligence models needs to be both effective and powerful in order to succeed with the integration of artificial intelligence and machine learning at edge devices. The optimization tools require perfect alignment with edge artificial intelligence because they combine quantization precision reduction tools alongside pruning methods for network connection removal. Model size decreases when floating-point values get quantized into 8-bit integer values from 32-bits. This greatly simplifies the calculations. The facial recognition tasks on real-time operate instantly through TensorFlow Lite framework when combined with Raspberry Pi and low-power platforms [7].

Federated learning improves The Edge Intelligence system for several models can simultaneously be trained instead of sharing heterogeneous type of raw data because Privacy issues depend on this. Recent implementations of adaptive federated learning have been deployed within IoT networks because of these platforms' limited resources capability. The official adaptive federated learning framework establishes an efficient balance, which enables more communication, while addresses existing resource limitations. The convergence method through federated learning enables better model accuracy and performance speeds than traditional practices. The Gboard keyboard from Google demonstrates how user prediction functions privately on individual devices. Furthermore, the healthcare federated applications perform diabetes diagnoses at a high level of confidentiality [8]. The performance and effectiveness of well-known lightweight AI optimization methods are presented in Table 1.

TABLE I
COMPARISON OF AI MODEL OPTIMIZATION TECHNIQUES AT THE EDGE[7-9]

Optimization Technique	Accuracy Impact	Model Size Reduction	Common Platforms
Model Pruning	Approximately -2%	40–50%	MobileNet, ResNet
Quantization	Approximately -1–3%	75–80%	TensorFlow Lite, ARM Cortex-M

II. APPLICATIONS OF EDGE INTELLIGENCE

The many advantageous applications of edge intelligence can be found across different sectors. Privacy, response and efficiency now distinguish these systems.

A. Smart Cities

Edge intelligence technology functions as an appropriate solution for managing smart city applications that need minimum delay times. The Barcelona traffic control systems operate artificial intelligence analytics through edge-based technology in various local points to achieve rapid processing of traffic camera information. Emergency response times increase as compared to the cloud-based systems. The delay in its responses decrease by 25% through the deployment of this technology. The deployment leads to reduced costs for network bandwidth while also providing financial operational efficiency. The fusion of AI drones working with edge computation today shows how to quickly detect fire anomalies in minimum time for early wildfire alerts through a method that avoids slow traditional cloud processing delays [11].

B. Healthcare

Edge intelligence brings essential functional advantages to real-time medical tracking diagnosis procedures. The Apple Watch Series 9 conducts ECG evaluations with wearable technology via its connection to localized machine learning algorithms. This system enables protected user privacy and generates 98% precise medical diagnoses [12]. Several protected patient databases can analyze medical knowledge be- cause medical models learn

collaboratively through federated learning. Predictive models receive modern psychological implementations that enable their deployment in diabetes systems management. The aggregation of multiple database data under secure conditions through federated learning creates better diagnostic accuracy and improved health results for patients [13].

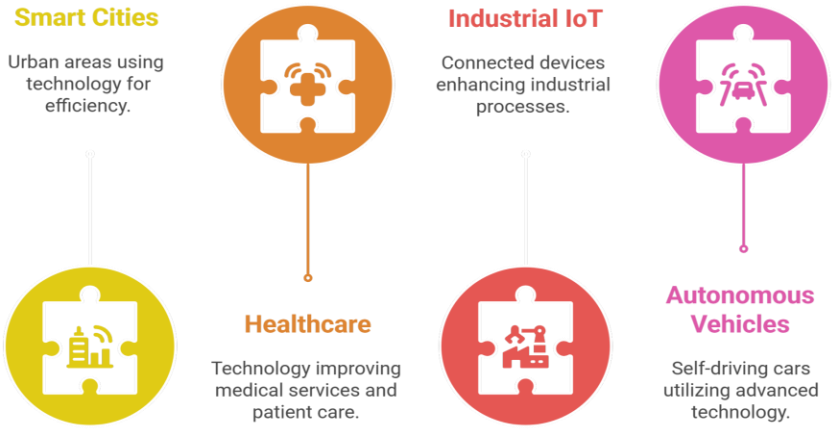

Fig. 2. Applications of Edge Intelligence

C. Industrial IoT (IIoT)

IIoT technology obtains its functionality from edge intelligence to support predictive maintenance systems and autonomous robots that deliver improved industrial output results. The financial circumstances of organizations improve substantially through edge-powered AI systems which ex- amine sensor information in real time and limit equipment downtime and maintenance expenses when doing predictive maintenance. Silicon-based predictive analysis from Siemens enables production equipment to prevent downtime. Thus generating $200000 worth of annual savings per line through its successful 40% avoidance rate. Amazon Kiva robots optimize their performance with autonomous navigation systems using LiDAR capabilities and edge-computing sensors and processing which lowers the need for human operators [15].

D. Autonomous Vehicles

The absolute requirement of edge intelligence exists for driverless car collision avoidance systems to ensure safety. The NVIDIA Drive AGX device from NVIDIA operates sensor data at a rate of 10 milliseconds. The system contains radar and LiDAR sensors and cameras among its components. Driver safety together with response requires the focused processing output. The system outperforms standard cloud systems because it enables processing that is faster than their 200- millisecond delays of cloud-based system, resulting potentially dangerous driving conditions [16].

Table 2: Edge Intelligence produces a comprehensive summary which compares its effects throughout various applications.

TABLE II
COMPARATIVE IMPACT OF EDGE INTELLIGENCE ACROSS KEY SECTORS[10-16]

Sector	Key Application	Latency Improvement	Bandwidth Savings	Cost/Operational Savings
Smart Cities	Real-time Traffic Management	25% improvement	40% reduction	97% reduction in operational costs
Healthcare	ECG & Diabetes Diagnostics	Real-time responses	99% data Privacy enhancement	Enhanced diagnostic accuracy

Industrial IoT	Predictive Maintenance	Significant downtime reduction	-	Approx. $200,000/year per production line
Autonomous Vehicles	Collision Avoidance	Less than 10 ms latency	-	Improved safety by avoiding significant travel distance during critical events

III. CHALLENGES AND SOLUTIONS

Edge intelligence increases IoT system capabilities while multiple technological as well as security and scalability problems persist unresolved. Global experts and business entities maintain their research into breakthrough solutions that address ongoing difficulties.

A. Technical Challenges

1. **Energy Consumption:** The ongoing main problem relates to energy use restrictions in IoT devices that have affect automated systems that need continuous operation and require either remote power supplies or built-in batteries. Artificial intelligence systems developed today need considerable processing capacity, yet their total capacity often approaches maximum usage. Alirozai, in his paper, has discussed that enhancing energy efficiency has reduced from Google's Coral Edge TPU, as it has lightweight hardware architectures. The power requirements of standard GPUs exceed 70-to-80 watts, while the recent examples of TPU shows that they consume electricity at rates between 2–5 watts. Spiking Neural Net- works facilitate large energy savings up to 80% as compared to Convolutional Neural Network (CNN)[18].

2. Latency-Accuracy Trade-off: The challenge for Edge Intelligence faces major obstacles when achieving high precision and low latency at the same time. The accuracy of models becomes reduced after applying techniques like model pruning together with quantization which also reduces model size. Cloud-Edge hybrid systems exist as a solution to this problem. The system performs exact assessments through cloud processing and delivers prompt results using edge computations. These methods help reduce delay requirements and provide an effective way to maintain suitable accuracy performance [18]. Table 3: The paper outlines important system and technologies together with their performance in resolving these significant technical problems:

TABLE III
SUMMARY OF TECHNICAL SOLUTIONS TO EDGE INTELLIGENCE CHALLENGES[17-20]

Technical Challenge	Solution Tech-nology	Key Advantage	Notable Application Platforms
Energy Efficiency	Edge TPUs (Coral)	Low power (2-5 W)	IoT devices, real- time inference
-	Spiking Neural Networks (SNNs)	80% power reduction	Bio-inspired computing, IoT sensors
Latency- Accuracy	Hybrid Edge-Cloud Architectures	Balanced accuracy and latency	Autonomous vehicles, smart healthcare

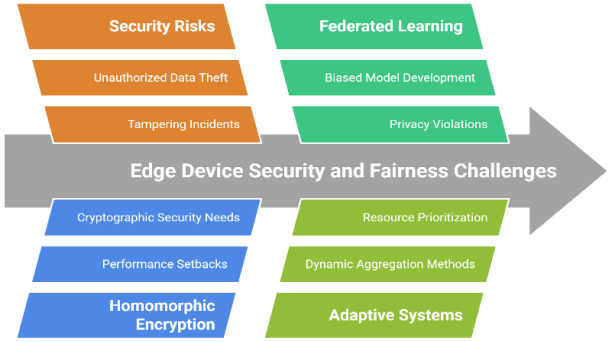

Fig. 3. Security and Fairness Challenges

B. Security and Fairness Challenges

Edge devices experience higher risks of tampering incidents along with unauthorized data theft due to their exposed positioning. The edge environment requires specific adaptations of homomorphic encryption (HE), which addresses cryptographic security needs. The ability to operate encrypted data at the edge without requiring decryption during the process leads to better security protection without significant performance setbacks [19]. FL prevents data centralization yet faces two main issues which include biased model development from localized information shortcomings and privacy violations of processed data. The solution of these problems become possible through dynamic aggregation method modifications in adaptive federated learning systems. The system architecture guarantees both fairness and security of analytics across all edge devices when minimizing diverse data conditions. Recent adaptive federated learning systems determine which parts of the model to prioritize through their decision elements based on available resources and data types. This approach decreases discrimination between system users.

C. Scalability Challenges

IoT system scalability remains complex because IoT incorporates numerous devices while manufacturers do not know precisely how long they will operate. The practical solution of edge computing has flexible frameworks, smart management systems and dynamic resource handling. Thus, these systems alter their operations in real-time for dependable performance during peak network conditions. With these designs, it is shown that they exhibit consistent performance, adaptability and dependability regardless of the environment [21].

IV. Future Trends

Several new developments are ready to improve Edge Intelligence's usability and appeal as technology and networks always change throughout IoT environments.

A. 6G Networks

The Six-generation (6G) network system changes every- thing by delivering quick data speed and solid system process- ing with rapid reply times through Edge Intelligence. The next six-generation network system will unite small-scale latency standards with 1 Tbps speed rates while servicing ten million devices in each square kilometer. Consumer electronics have reached exceptional speed milestones which boost latency- sensitive applications starting from robotic surgery up to virtual reality and driverless automobiles [22,23].

Table 4: The comprehensive outline depicts major improvements which 6G will bring to 5G technology standards.

TABLE IV

Comparative Overview of 5G and Projected 6G Network Capabilities[22]

Metric	5G (Current Standard)	6G (Projected Standard)
Latency	1–10 ms	less than 0.1 ms
Peak Data Rate	Up to 10 Gbps	Up to 1 Tbps
Device Density	1M devices/km²	10M devices/km²
Reliability	99.999%	99.99999%

B. Ethical AI and Regulatory Compliance

Ethical matters together with rule compliance will shape how artificial intelligence can be deployed at the edge. Ac- cording to the European Union's AI Act all regulatory bench- marks regarding fairness alongside openness responsibility and prejudice avoidance are established. These edge intelligence systems should maintain proximity to the specified models. The scientific community needs to make explainable artificial intelligence together with open model predictions and bias reduction their top priority during research and development activities. The rules mentioned above specifically protect sensitive applications that serve law enforcement combined, public safety, healthcare sectors and other fields, which emphasize equity and responsibility [24].

C. Blockchain Integration

The implementation of blockchain technology serves as an obvious trend because it provides secure distributed administration and integrity to edge settings. Edge Intelligence working with blockchain technology permits IoT devices to manage secure transactions and data operations through distributed systems that lack central authorities. Researches have established that blockchain serves as an efficient mechanism to safeguard distributed IoT networks particularly for private transactions, data integrity, and distributed trust management [25].

V. CONCLUSION

Current traditional cloud based solutions have a number of limitation without having any solution to date, which cannot solve most of the IoT problems. More and more, edge devices are being made more efficient and more secure as a technology by going through model pruning, quantization and federated learning. We are furthermore living in an era of edge intelligence with hybrid aircraft, blockchain technology and incoming 6G networks which will make the factors we have discussed here a bit more higher. Both sectors have a very high scope of impact in areas including healthcare, smart cities and autonomous vehicles. The journey is arduous, though. However, major barriers include energy inefficiency and security vulnerabilities and latency-accuracy barrier. And all of these things have to be done by research, policymakers and industry stakeholders together. Such a project will focus on the areas that allow for lightweight AI models, ethical AI frameworks and the scaling of edge solutions to power billions of devices. The road ahead demands continuous innovation and ethical responsibility. With synergistic progress of AI and communication technologies, IoT can be transformed to more intelligent, rapid and autonomous systems for a better tomorrow.

Siemens. (2024). Edge AI for Predictive Maintenance in Industrial IoT. Siemens Industrial Solutions Whitepaper.

[14] Amazon Robotics. (2023). Edge Computing for Autonomous Warehouse Robots Navigation. Amazon Research Reports.

[15] NVIDIA Corporation. (2024). NVIDIA Drive AGX Platform: Real-time Edge Processing for Autonomous Vehicles. IEEE Intelligent Transportation Systems Magazine, 16(1), 56-63.

[16] Google Coral. (2024). Edge TPU: AI Accelerator for Ultra-Low Power IoT. Google AI Research Reports.

[17] Zhang, Q., Li, F., Wang, X., Liu, H. (2023). Balancing Latency and Accuracy in Edge Computing Environments. IEEE Transactions on Parallel and Distributed Systems, 34(3), 1251-1262.

[18] Wang, J., Xu, C., Luo, M., Yang, S. (2023). Optimizing Homomorphic Encryption for Privacy-Preserving Edge Computing. IEEE Internet of Things Journal, 10(7), 5298-5309.

[19] Chen, Y., Liu, H., Zhou, J., Wu, Y. (2024). Adaptive Federated Learning to Mitigate Bias in Edge AI. ACM Transactions on Intelligent Systems and Technology, 15(1), 25-48.

[20] Liu, M., Zhang, W., Li, Q. (2023). Scalable Edge Intelligence: Managing Dynamic Workloads in Heterogeneous IoT Systems. IEEE Communications Magazine, 61(11), 48-54.

[21] Liu, G., Wang, X., Huang, S., Chen, Z. (2024). Edge Intelligence in 6G Networks: Unlocking New Capabilities for IoT. IEEE Communications Magazine, 62(1), 48-55.

[22] Kumar, S., Roy, S., Patel, D. (2023). Digital Twins in Industrial IoT: Applications and Benefits for Predictive Analytics. IEEE Industrial Electronics Magazine, 17(2), 45-53.

[23] European Commission. (2024). EU AI Act: Regulatory Framework for Artificial Intelligence. Official Journal of the European Union.

[24] Sarker, S., Ma, J., Park, J. (2025). Blockchain and Edge Intelligence for Secure IoT: A Comprehensive Review. IEEE Communications Surveys Tutorials, 27(1), 153-174.

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.

[2] 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.

[3] Ahmed, E., Rehmani, M. H., Yaqoob, I. (2017). Edge Computing in IoT: A Survey. IEEE Communications Surveys Tutorials, 19(4), 2322- 2358.

[4] Konecny´, J., McMahan, H. B., Ramage, D., Richtarik, P. (2016). Federated Learning: Strategies for Improving Communication Efficiency. arXiv preprint arXiv:1610.05492.

[5] McKinsey Company. (2022). IoT and Edge Computing: Capturing the Value at Speed. McKinsey Digital Report.

[6] Gartner. (2023). Top Strategic Technology Trends for Edge Computing. Gartner Research Report.

[7] Han, S., Mao, H., Dally, W. J. (2015). Deep Compression: Compress- ing Deep Neural Networks with Pruning, Trained Quantization, and Huffman Coding. International Conference on Learning Representations (ICLR), 2015.

[8] Li, H., Chen, X., Wang, Z., Li, Y., Tian, Y. (2024). Adaptive Federated Learning for Resource-Constrained IoT Devices in Edge Computing. Scientific Reports, 14(1).

[9] Zhou, Y., Zhao, J., Zhang, X. (2023). Edge Intelligence: Edge Com- puting for 5G and the Internet of Things. Future Internet, 17(3), 101.

[10] Ahmed, E., Rehmani, M. H., Yaqoob, I. (2017). Edge Computing in IoT: A Survey. IEEE Communications Surveys Tutorials, 19(4), 2322- 2358.

[11] Akbari, Y., Hasan, M., Khan, S. (2023). Edge AI-based Drone Surveillance for Wildfire Detection. IEEE Transactions on Industrial Informatics, 19(8), 9485-9493.

[12] Apple Inc. (2024). Apple Watch Series 8: ECG App Accuracy White Paper. Apple Research Reports.

[13] Zhu, L., Liu, Y., Han, X., Hu, G. (2023). Federated Learning for Diabetes Prediction: A Privacy-preserving Approach. MDPI Sensors, 23(8), 4560..