IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

# Cloud-Assisted Edge AI: Enhancing Decision Making in IoT Devices with Cloud-Powered Machine Learning Models

**Urvi C. Gupta, Roshni P. Adiyecha, Hitesh A. Solanki**

Department of computer Science, Sacred Heart College, Tirupattur, S.India

**ABSTRACT:** The Internet of Things (IoT) is transforming industries by enabling devices to gather and share data. However, IoT devices often face limitations in processing power, storage, and energy consumption, restricting their ability to make complex decisions in real time. To address these challenges, cloud-assisted edge AI combines the advantages of edge computing and cloud-powered machine learning models, enabling IoT devices to make intelligent decisions at the edge while leveraging cloud resources for more complex processing tasks. This paper explores the integration of edge AI with cloud assistance, demonstrating how this hybrid approach enhances decision-making capabilities in IoT devices. The paper examines key concepts, the architecture of cloud-assisted edge AI systems, application areas, and the benefits of combining edge and cloud computing. Furthermore, challenges such as latency, data privacy, and integration issues are discussed, alongside future directions for this technology in IoT applications.

**KEYWORDS**: Cloud-Assisted Edge AI, IoT, Machine Learning, Edge Computing, Decision Making, Cloud Computing, Real-Time Processing, Artificial Intelligence, Data Privacy, Latency

## I. INTRODUCTION

The rapid growth of the Internet of Things (IoT) has led to an increasing number of connected devices, ranging from smart home appliances to industrial machines. These devices generate large volumes of data that can be used for insights, optimization, and automation. However, many IoT devices are constrained in terms of computational power, storage, and energy efficiency. As a result, processing large amounts of data locally (at the device level) can be challenging, especially when real-time decision-making is required.

Edge AI addresses this issue by bringing machine learning (ML) models closer to the source of the data, enabling devices to make intelligent decisions locally. However, edge devices may still lack the necessary computational resources to process complex models. Cloud-assisted edge AI solves this problem by offloading heavy computational tasks to the cloud while allowing edge devices to execute simpler models and make real-time decisions. This hybrid

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

approach improves the performance of IoT systems by enabling both real-time decision-making and complex data analysis.

## II.ARCHITECTURE OF CLOUD-ASSISTED EDGE AI

The architecture of a cloud-assisted edge AI system integrates edge computing and cloud computing to optimize decision-making processes for IoT devices. The system typically consists of the following components:

### 2.1 Edge Devices

Edge devices are the first point of interaction with the environment. These devices are responsible for collecting sensor data, performing local data processing, and making decisions using lightweight machine learning models. Examples of edge devices include sensors in smart cities, autonomous vehicles, and wearables.

### 2.2 Edge AI Models

Edge AI models are lightweight machine learning algorithms deployed directly on IoT devices. These models are designed to process data in real time with limited resources. Examples of edge AI models include decision trees, regression models, and lightweight neural networks. These models help edge devices make initial decisions, such as detecting anomalies or triggering alerts.

### 2.3 Cloud Computing Infrastructure

The cloud provides the computational resources necessary for training complex machine learning models. In a cloud-assisted edge AI system, the cloud serves as the central repository for large datasets and performs heavy-duty processing tasks, including model training, optimization, and advanced analytics.

### 2.4 Data Communication Layer

The communication layer enables secure and efficient data transfer between edge devices and the cloud. IoT devices periodically send data to the cloud for advanced processing, and cloud-trained models are sent back to the edge for execution. This layer also ensures that data privacy and security concerns are addressed through encryption and secure protocols.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

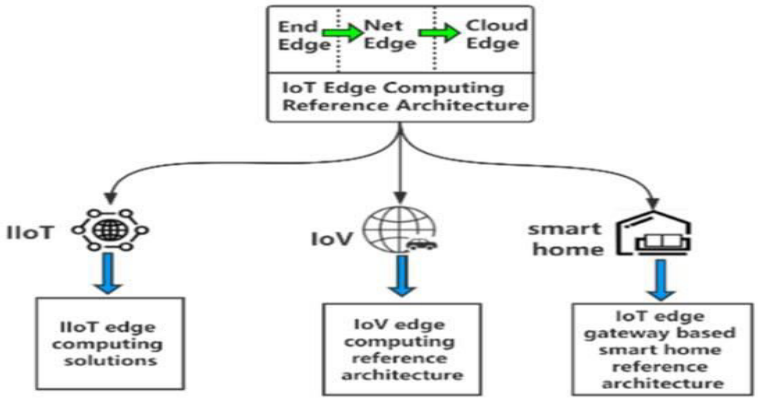
Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

**Figure 1**: Architecture of Cloud-Assisted Edge AI for IoT Devices




## III.APPLICATIONS OF CLOUD-ASSISTED EDGE AI IN IOT

Cloud-assisted edge AI is applicable in a wide range of industries and applications, where real-time decision-making and advanced machine learning are needed. Some key application areas include:

### 3.1 Smart Cities

In smart cities, IoT devices such as traffic cameras, environmental sensors, and waste management systems generate large amounts of data. Cloud-assisted edge AI can help optimize traffic flow, improve public safety, and reduce energy consumption by processing data locally for immediate decision-making and sending complex analytics to the cloud.

- **Traffic Management**: Edge AI models can detect traffic patterns and adjust traffic signals in real time, while the cloud analyzes traffic data for long-term trends and urban planning.
- **Public Safety**: Edge AI can detect unusual behavior or events in real time (e.g., accidents or security breaches), while the cloud can perform facial recognition and other advanced analysis.

### 3.2 Healthcare and Remote Monitoring

IoT devices in healthcare, such as wearable devices or smart medical instruments, monitor patient vitals and detect anomalies in real-time. By combining edge AI with cloud processing, these devices can make immediate recommendations (e.g., alerting medical staff in case of emergencies) while relying on the cloud to train more advanced models for personalized care.

- **Real-Time Health Monitoring**: Wearable devices can use edge AI to detect abnormal patterns (e.g., irregular heartbeats) and notify healthcare providers immediately.
- **Personalized Healthcare**: Cloud computing enables the processing of large datasets from multiple sources to develop personalized treatment plans based on advanced machine learning models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

### 3.3 Industrial IoT (IIoT) and Predictive Maintenance

In industrial environments, IoT sensors monitor machinery, environmental conditions, and production processes. Cloud-assisted edge AI helps in predictive maintenance by enabling devices to make real-time decisions about machine health while offloading complex predictive models to the cloud for continuous improvement.

- **Predictive Maintenance**: Edge devices analyze sensor data in real time to detect early signs of failure, while the cloud performs long-term trend analysis and model refinement.
- **Production Optimization**: Real-time data from IoT devices can help optimize production lines, reduce downtime, and improve energy efficiency.

### 3.4 Autonomous Vehicles

Autonomous vehicles rely on IoT sensors (such as cameras, LIDAR, and radar) to make real-time decisions on navigation, object detection, and collision avoidance. Cloud-assisted edge AI enables the vehicle to process data locally for immediate decisions while relying on the cloud for advanced analytics, such as route optimization and fleet management.

## IV.BENEFITS OF CLOUD-ASSISTED EDGE AI

The combination of edge computing and cloud resources offers several advantages for IoT applications:

### 4.1 Low Latency and Real-Time Decision Making

By processing data at the edge, cloud-assisted edge AI enables faster decision-making. This is critical in applications like autonomous vehicles, smart cities, and healthcare, where real-time responses are required.

### 4.2 Reduced Bandwidth and Energy Consumption

Edge devices can process data locally and only send necessary information to the cloud, reducing the amount of data transmitted and conserving bandwidth and energy. This is particularly beneficial in remote areas where connectivity is limited.

### 4.3 Scalability

Cloud infrastructure allows for scalable data storage, processing power, and machine learning model training. As the number of IoT devices increases, cloud-assisted edge AI can easily scale to handle the growing data.

### 4.4 Improved Security and Privacy

By processing sensitive data locally at the edge, cloud-assisted edge AI enhances data privacy. Only aggregated data or critical insights are sent to the cloud, reducing the risk of exposing sensitive personal or industrial data.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

**Table 1**: Benefits of Cloud-Assisted Edge AI

| Benefit | Description |
|---|---|
| **Low Latency** | Real-time decision-making with local processing. |
| **Reduced Bandwidth Usage** | Only necessary data is sent to the cloud, conserving bandwidth. |
| **Energy Efficiency** | Local processing reduces energy consumption compared to cloud-only systems. |
| **Scalability** | Cloud infrastructure scales with increasing IoT devices and data. |
| **Enhanced Security** | Sensitive data is processed locally, ensuring better privacy. |

## V.CHALLENGES AND FUTURE DIRECTIONS

While cloud-assisted edge AI provides significant advantages, there are challenges that must be addressed for successful deployment:

### 5.1 Latency and Data Synchronization

Despite edge processing, communication between edge devices and the cloud can introduce latency. This may hinder applications that require instantaneous decision-making.

### 5.2 Data Security and Privacy

Handling sensitive data in both cloud and edge environments poses privacy risks. Ensuring secure communication between devices and the cloud, as well as complying with data privacy regulations, is a challenge.

### 5.3 Complexity in Integration

Integrating edge computing, machine learning models, and cloud infrastructure can be complex, especially when using devices from multiple manufacturers or when transitioning legacy systems to cloud-assisted edge AI.

### 5.4 Computational Constraints at the Edge

While edge devices are increasingly capable, they still have limited computational power compared to cloud systems. Designing efficient algorithms and models that balance edge capabilities with cloud assistance is key.

## VI.CONCLUSION

Cloud-assisted edge AI provides a powerful framework for enhancing decision-making capabilities in IoT devices by combining real-time edge computing with cloud-powered machine learning models. This hybrid approach offers numerous benefits, including low latency, reduced energy consumption, scalability, and improved security. As the number of IoT devices continues to grow, cloud-assisted edge AI will play a crucial role in optimizing applications

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

across industries, from smart cities to healthcare to autonomous vehicles. However, addressing challenges related to latency, security, and integration will be essential for realizing the full potential of this technology.

## REFERENCES

1. Zhang, Y., & Yang, X. (2021). *Edge AI for IoT: Architectures and Applications*. Springer.
2. Li, W., & Liu, L. (2020). *Cloud-Edge Intelligence for IoT: Models, Algorithms, and Applications*. IEEE Transactions on Industrial Informatics, 16(4), 2481-2491.
3. Sivathapandi P, Sudharsanam SR, Manivannan P. Development of Adaptive Machine Learning-Based Testing Strategies for Dynamic Microservices Performance Optimization. Journal of Science & Technology. 2023 Mar 21;4(2):102-37.
4. Sugumar, Rajendran (2023). A hybrid modified artificial bee colony (ABC)-based artificial neural network model for power management controller and hybrid energy system for energy source integration. Engineering Proceedings 59 (35):1-12.
5. G Jaikrishna, Sugumar Rajendran, Cost-effective privacy preserving of intermediate data using group search optimisation algorithm, International Journal of Business Information Systems, Volume 35, Issue 2, September 2020, pp.132-151.
6. Sugumar R., et.al IMPROVED PARTICLE SWARM OPTIMIZATION WITH DEEP LEARNING-BASED MUNICIPAL SOLID WASTE MANAGEMENT IN SMART CITIES, Revista de Gestao Social e Ambiental, V-17, I-4, 2023.
7. Akash, T. R., Lessard, N. D. J., Reza, N. R., & Islam, M. S. (2024). Investigating Methods to Enhance Data Privacy in Business, Especially in sectors like Analytics and Finance. Journal of Computer Science and Technology Studies, 6(5), 143–151.https://doi.org/10.32996/jcsts.2024.6.5.12
8. Seethala, S. C. (2024). AI-Infused Data Warehousing: Redefining Data Governance in the Finance Industry. International Research Journal of Innovations in Engineering & Technology, 5(5), Article 028. https://doi.org/10.47001/IRJIET/2021.505028
9. Miao, Y., & Li, C. (2022). *Hybrid Cloud-Edge Computing for IoT Applications: Challenges and Solutions*. Elsevier Journal of Computer Networks, 183, 107466.
10. S. Muthubalaji, Archana Saxena (2024). The Structured use of ML Technique in Creation of Powerful 7-D based Gaming Tools. International Conference on Advance Computing and Innovative Technologies in Engineering 4 (1):1263-1267.
11. Amazon Web Services (AWS). (2020). *AWS IoT and Edge Computing for Real-Time Decision Making*. AWS Whitepapers. Retrieved from https://aws.amazon.com
12. Tarun Prashar, Sandeep Kumar (2024). Distribution Carried Automation System via Radical Substantial strap Technology. International Conference on Advance Computing and Innovative Technologies in Engineering 4 (1):1322-1326.
13. Vimal Raja, Gopinathan (2024). Intelligent Data Transition in Automotive Manufacturing Systems Using Machine Learning. International Journal of Multidisciplinary and Scientific Emerging Research 12 (2):515-518.
14. Muntather Almusawi, Harpreet S. Bhatia (2024). The Structured Design Framework for Developing Discharging Strategy for Cloud Based Automation Through ML Technique. International Conference on Advance Computing and Innovative Technologies in Engineering 4 (1):1341-1345.
15. Karandikar, A.S. (2024). Cybersecurity in Telecom: Protecting Software Systems in the Digital Age. International Journal of Computer Engineering and Technology (IJCET), 15(5), 658–665.
16. Microsoft Azure. (2021). *Intelligent Edge: Combining AI and Cloud for IoT Applications*. Microsoft Azure Blog. Retrieved from https://azure.microsoft.com
17. Praveen Sivathapandi, Prabhu Krishnaswamy (2022). Advanced AI Algorithms for Automating Data Preprocessing in Healthcare: Optimizing Data Quality and Reducing Processing Time. Journal of Science and Technology (Jst) 3 (4):126-167.
18. Vimal Raja, Gopinathan (2022). Leveraging Machine Learning for Real-Time Short-Term Snowfall Forecasting Using MultiSource Atmospheric and Terrain Data Integration. International Journal of Multidisciplinary Research in Science, Engineering and Technology 5 (8):1336-1339.
19. Arulraj AM, Sugumar, R., Estimating social distance in public places for COVID-19 protocol using region CNN, Indonesian Journal of Electrical Engineering and Computer Science, 30(1), pp.414-424, April 2023.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

20. Sugumar, R., Rengarajan, A. & Jayakumar, C. Trust based authentication technique for cluster based vehicular ad hoc networks (VANET). Wireless Netw 24, 373–382 (2018). https://doi.org/10.1007/s11276-016-1336-6
21. Prasad, G. L. V., Nalini, T., & Sugumar, R. (2018). Mobility aware MAC protocol for providing energy efficiency and stability in mobile WSN. International Journal of Networking and Virtual Organisations, 18(3), 183-195.
22. Megha Pandey, Subramani K. (2024). An Innovative Way of Trackable GDS in the Field of CC. International Conference on Advance Computing and Innovative Technologies in Engineering 4 (1):1
23. Dynamic Interactive Multimodal Speech (DIMS) Framework. (2023). Frontiers in Global Health Sciences, 2(1), 1-13. https://doi.org/10.70560/1s1ky152
24. Deepak Kumar, Laith H. Alzubaidi (2024). The Different Way of Utilizing the Intellectual of Artificial Intelligence in the Animal Farming Field Progress of AI. International Conference on Advance Computing and Innovative Technologies in Engineering 4 (1):1624-1626.
25. Kartheek, Pamarthi (2023). Big Data Analytics on data with the growing telecommunication market in a Distributed Computing Environment. North American Journal of Engineering and Research 4 (2).[11]
26. Vaka, P. R. (2021). Zero Trust Security Model. International Journal of Advanced Research in Engineering and Technology (IJARET), 12(6), 148–156.
27. Arul Raj .A.M and Sugumar R.," Monitoring of the social Distance between Passengers in Real-time through video Analytics and Deep learning in Railway stations for Developing highest Efficiency" , March 2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2022, ISBN 979- 835033384-8, March 2023, Chennai , India ., DOI 10.1109/ICDSAAI55433.2022.10028930.
28. Soundappan, S.J., Sugumar, R.: Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. Int. J. Bus. Intell. Data Min. 11, 338 (2016)
29. Sugumar, R. (2023). Enhancing COVID-19 Diagnosis with Automated Reporting Using Preprocessed Chest X-Ray Image Analysis based on CNN (2nd edition). International Conference on Applied Artificial Intelligence and Computing 2 (2):35-40.
30. K. Thandapani and S. Rajendran, "Krill Based Optimal High Utility Item Selector (OHUIS) for Privacy Preserving Hiding Maximum Utility Item Sets", International Journal of Intelligent Engineering & Systems, Vol. 10, No. 6, 2017, doi: 10.22266/ijies2017.1231.17.
31. Sugumar, R. (2023). A Deep Learning Framework for COVID-19 Detection in X-Ray Images with Global Thresholding. IEEE 1 (2):1-6.
32. P. Manjula, K. Krishnakumar (2024). A Novel Method for Detecting Liver Tumors combining Machine Learning with Medical Imaging in CT Scans using ResUNet. International Conference on Integrated Circuits and Communication Systems 1 (1):1-5.
33. P. Pulivarthy. "Enhancing data integration in oracle databases: Leveraging machine learning for automated data cleansing, transformation, and enrichment". International Journal of Holistic Management Perspectives, vol. 4, no. 4, pp. 1-18, 2023.
34. Vikram A., Ammar Hameed Shnain (2024). AI-Powered Network Intrusion Detection Systems. International Conference on Communication, Computing and Signal Processing 1 (1):1-6.
35. Kartheek, Pamarthi (2023). Protecting the Hadoop Cluster on the Basis of Big Data Security. Journal of Artificial Intelligence, Machine Learning and Data Science 1 (3):831-837.
36. Lokesh Kalapala, D. Shyam (2024). Research on Reasonable Color Matching Method of Interior Decoration Materials Based on Image Segmentation. International Conference on Smart Technologies for Smart Nation 2 (1):1001-1006.
37. Jose N. N., Deipali Gore (2024). Efficient predefined time adaptive neural network for motor execution EEG signal classification based brain-computer interaction. Elsevier 1 (1):1-11.
38. Kartheek, Pamarthi (2022). Applications of Big Data Analytics for Large-Scale Wireless Networks. Journal of Artificial Intelligence, Machine Learning and Data Science 1 (1):920-926.
39. Sugumar, R. (2022). Estimation of Social Distance for COVID19 Prevention using K-Nearest Neighbor Algorithm through deep learning. IEEE 2 (2):1-6.
40. R. Sugumar, A. Rengarajan and C. Jayakumar, Design a Weight Based Sorting Distortion Algorithm for Privacy Preserving Data Mining, Middle-East Journal of Scientific Research 23 (3): 405-412, 2015.
41. Dong Wang, Lihua Dai (2022). Vibration signal diagnosis and conditional health monitoring of motor used in biomedical applications using Internet of Things environment. Journal of Engineering 5 (6):1-9.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312206|

42. K. KrishnaKumar, M. Jenifer Pallavi M. Shanthappa (2024). Molecular insights into the structural, spectroscopic, chemical shift characteristics, and molecular docking analysis of the carbamate insecticide fenobucarb. Elsevier 1 (1):1-12.
43. Praveen Sivathapandi, Girish Wali (2023). MULTI AGENT MODEL BASED RISK PREDICTION IN BANKING TRANSACTION USING DEEP LEARNING MODEL. JOURNAl OF CRITICAL REVIEWS 10 (2):289-298.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details