Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 4 (2025)
https://www.jneonatalsurg.com

Comparing and Selecting The Performance Efficiency Of Computing Techniques For Real-Time Image Processing

Won-hyuk Choi[1], Woo-Jin Jung[*2]

[1]Department of Avionics, Hanseo University

[2]Department of Aeronautical System Engineering, Hanseo University

Cite this paper as: Won-hyuk Choi, Woo-Jin Jung, (2025) Comparing and Selecting The Performance Efficiency Of Computing Techniques For Real-Time Image Processing. *Journal of Neonatal Surgery*, 14 (4), 284-290.

ABSTRACT

Recently, increase in the importance of real-time data processing, it is essential to choose an efficient computing method. This study finds the optimal computing method by comparing the performance of cloud computing (CC), edge computing (EC), and edge computing (ODEC) using on-device edge computing (ODEC). In particular, we compare the strengths and weaknesses of each technology by taking examples in areas where fast data processing and real-time response are important, such as real-time video processing. According to the research results, it can be seen that cloud computing greatly increases latency due to bottlenecks in data transmission, while edge computing and on-device AI technology can minimize latency thanks to distributed structures. It compares the performance of each technology at various data scales and emphasizes that on-device AI-based approaches perform well in environments that are less affected by the network and require large-capacity data processing and real-time response. This presents the possibility of overcoming the limitations of existing computing models and developing into smarter systems.

Keywords: *Cloud Computing, Edge Computing, AI-based On-device Edge Computing, Real-time Processing, Data Analysis.*

1. INTRODUCTION

In this study, Recently, as the amount of digital data has increased exponentially, the importance of data processing and management is increasing. Against this background, data processing systems built with cloud computing, edge computing, and lightweight AI edge computing have become basic requirements. Among them, cloud computing, edge computing, and lightweight AI edge computing are considered leaders. Cloud computing is widely used in environments that require large-scale data processing and efficient resource management, such as corporate business processing and online service provision. Recently, 'Cloud AI', which enables faster and more accurate data processing in combination with AI technology, has been applied in various fields. Using cloud AI, companies have the advantage of making faster and more accurate decisions. Edge computing has the advantage of reducing network traffic generated in the connection with cloud computing by processing large-scale data locally without transferring it to the cloud, and enables fast decision-making by processing and analyzing large-scale data generated from IoT devices. Lightweight AI edge computing is a technology that can process and analyze data locally and is very useful in situations that require a quick response. These technologies are applied to various fields and can also be used in geographic and urgent medical and safety processing. This paper analyzes the difference between modeling algorithms and hardware use of the three technologies to establish an experimental environment for performance comparison and focuses on real-time large-capacity data processing situations such as corridor monitoring. Through this, we present which technology is most efficient to use in each situation, improve the understanding of each technology, and review its appropriate applicability. In addition, by comprehensively analyzing the performance and limitations of these technologies, we will identify the strengths and weaknesses of each technology and propose the technology that can be used most effectively in the optimal situation. 1Faculty of Social Sciences and Humanities, Putera Batam University, Indonesia. E-mail: mortigor.afrizal@gmail.com

2. COMPUTING DATA PROCESSING RESEARCH

2.1 Edge Computing

Edge Computing (EC) is a technology that processes large-scale data generated in an IoT environment at the boundary of a network to speed up data analysis and provide rapid services to users. In particular, the importance of edge computing is remarkable in situations such as corridor monitoring, which requires large amounts of data to be processed in real time.

Corridor monitoring is one of the appropriate cases for the use of edge computing because it collects and analyzes large amounts of data in real time. When edge computing is applied, data collected from the corridor monitoring device can be processed immediately on the spot and quickly provide the information necessary for real-time decision making. This is effective in reducing delays that can occur by directly processing data collected from IoT devices on the spot, unlike cloud computing methods that process all data on a central server. In addition, the edge cloud can minimize service latency and efficiently utilize the resources of the central server by initially collecting and analyzing IoT data on the spot and then transmitting only the necessary information to the central cloud. This method can optimize the performance of IoT services and greatly improve the efficiency of data processing.

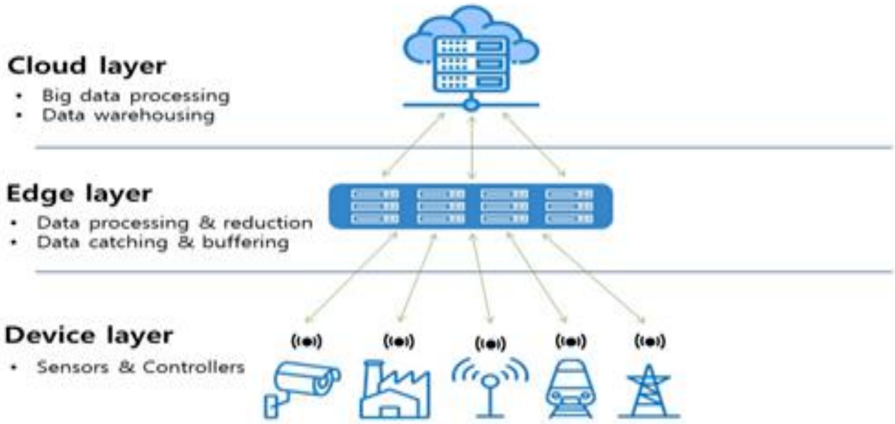

Figure 1.Road Freezing (Black Ice)

2.2 Cloud Computing

Cloud Computing (CC) provides better processing performance than edge devices, and the conversion of computing tasks to the cloud is widely recognized as a way to increase the efficiency of data processing. Cloud computing performs well in processing large amounts of data through a centralized structure and provides important advantages in situations where large-scale data such as corridor monitoring must be processed in real-time. This has the potential to increase the efficiency of enterprise IT resource management and improve data processing speed. Cloud computing is utilized not only in data storage and processing, but also in various industries and plays an important role in responding to continuous technological development and market demand. These technologies provide the possibility to fundamentally change the IT infrastructure of companies and individuals, and their importance is gradually emerging. However, the centralized structure of cloud computing can cause bottlenecks in the process of large-scale data transmission, which can cause problems such as service delays or interruptions. To solve this problem, this paper compares the utility of on-device AI computing (On-device AI Computing) technology. On-device AI minimizes delays and reduces bottlenecks that may occur during the data transmission process by processing in real time on the device itself without transferring data to the cloud. This technology can greatly improve stability, security, and performance, and provides a way to build a cooperative computing environment between the cloud and edge devices while complementing the shortcomings of cloud computing. This cooperative approach enables faster and more stable data processing while maintaining the scalability of cloud computing.

2.3 On-Device AI Edge Computing

On-Device AI-based intelligent edge computing (OEC)** is a technology that processes, analyzes, and transmits data by executing an AI model on the device itself, and has the characteristics of providing real-time responsiveness and large-scale data processing capabilities at the same time. In particular, this is a very advantageous technology in situations where large amounts of data have to be processed in real-time, such as real-time video processing. Corridor monitoring is a task of monitoring and predicting the marine environment in real-time, and is a representative example in which intelligent edge computing using on-device AI can be effectively applied. On-device AI has the advantage of being able to respond in real time without transferring data to an external server by processing data directly inside the device.

This method is suitable for rapid collection and analysis of large-scale data occurring in the ocean and provides the ability to process immediately at the device level. Local devices are responsible for real-time processing and initial analysis, and data centers support higher-order tasks such as more complex analysis and large-scale data storage. Cloud services complement complex operations that devices cannot handle and increase the efficiency of the entire system. By combining virtualization and artificial intelligence (AI) technologies, on-device AI can convert vast amounts of data collected from devices into valuable information in real time. In particular, AI algorithms can detect complex or unexpected phenomena more accurately and have the ability to analyze them in real-time. These characteristics allow on-device AI-based intelligent edge computing to greatly improve the accuracy and efficiency of real-time image processing tasks, and to quickly respond to emergency situations such as disasters by utilizing mobility and local processing power. This study analyzes the technical characteristics and performance of on-device AI-based intelligent edge computing to explore the development potential of smart systems that can overcome the limitations of cloud computing and traditional edge computing. The on-device AI-based intelligent edge computing architecture used for real-time processing is presented in Figure 2.

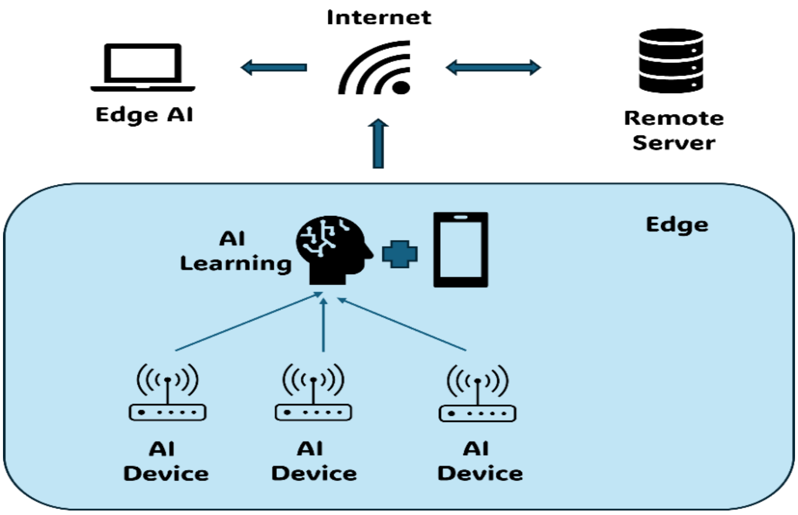

Figure 2. AI-based intelligent edge computing architecture

2-4 On-Device AI Edge Computing

The following is an equation for deriving the parameters used for the experiment

1. Cloud Computing

Parameters to calculate the local delay in cloud computing.

$$D_q = Queuing\ Delay$$

$$D_d = device\ Delay$$

$$D_b = \frac{Data\ size(byte)}{Bandwidth}$$

$$D_p = \frac{Distance}{Propagation\ Speed}$$

Total local delay formula for cloud computing.

$$D_c = D_p + D_b + D_d + D_q$$

Total added penalty local delay formula for cloud computing.

$$D_{bloss} = D_b * (1 - Packetloss)$$

$$D_c = D_p + D_{bloss} + D_d + D_q$$

2. Edge Computing

Parameters to calculate the local delay in edge computing.

$$D_b = \frac{Data\ size(byte)}{Bandwidth}$$

$$D_d = Device\ Delay$$

Total local delay formula for edge computing.

$$D_e = D_b + D_d$$

3. On-device AI Edge computing

Parameters to calculate the local delay in On-device AI edge computing.

$$D_b = \frac{Data\ size(byte) * 0.7}{Bandwidth}$$

$$D_d = Device\ Delay$$

Total local delay formula for On-device AI edge computing.

$$D_e = D_b + D_d$$

3. RESEARCH RESULTS

3.1 Temperature Simulation Results

In this study, various environments for implementation of cloud computing, edge computing, and on-device edge computing were set and compared. In Table 1, parameter values are set so that fair comparisons can be made between different computing models. In this way, you can see how effectively each computing model works in different environments and requirements. In the case of On-device AI Edge, the 30% compression rate that has undergone average data processing through AI was set as a parameter to differentiate it.

Table 1: parameter table

Parameters	Setting
Data Size, R	10000 MB
Bandwidth Cloud, B_c	800 Mbps
Bandwidth Edge, B_e	80 Mbps
Bandwidth AI Edge, B_a	80 Mbps
GPU Cloud, G_c	8000 GFLOPS
GPU Edge,AI Edge, G_e, G_a	1000 GFLOPS
CPU Cloud, C_c	1500 GFLOPS
CPU Edge,AI Edge, C_e, C_a	100 GFLOPS
RAM Cloud, F_c	256GB
RAM Edge,AI Edge, F_e, F_a	16GB
Distance Factor Cloud, d_c	0.08
Distance Factor Edge,AI d_e, d_a	0.016
compression AI Edge, c_a	30%
Infra penalty	0, 100, 200

3.2 Compute Processing Performance Comparison

The results for the average delay time can be seen in Figure 3, clearly showing the differences in performance between cloud computing, edge computing, and on-device AI-based intelligent edge computing. Considering additional infrastructure, networks, and multiprocessing situations, cloud computing has a relatively long delay time because it is not real-time processing at first, but it continues to maintain a consistently good value. Edge computing and on-device AI-based edge computing tend to increase linearly as the data size increases. This is because the above graph does not actually show efficiency in real-time situations because it is a graph when only hardware performance parameters and bandwidth are used. In particular, in the case of on-device AI-based intelligent edge computing, data is processed directly inside the device, so even if the data size increases, there is little delay time locally because it actually processes immediately.

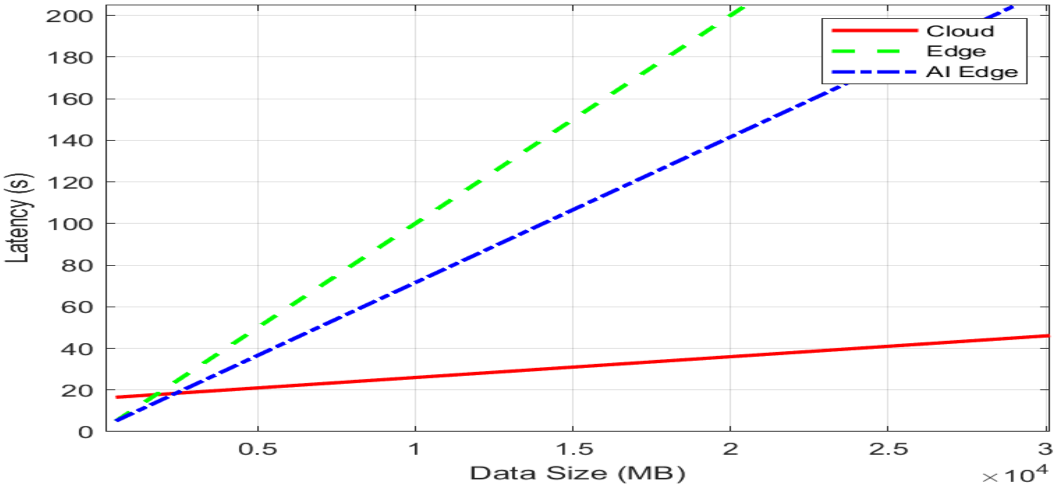

Figure 3. Average latency by data size

The results of the average processing time can be seen in Figure 4, which shows that cloud computing maintains relatively constant processing time despite changes in data size. This is because cloud servers have strong computing power and stable processing power, while the latency caused by data transfer does not work significantly. However, edge computing and on-device AI-based intelligent edge computing show patterns of shorter processing time as the data size increases. This is because the more local processing data there is, the more distributed processing power is maximized. In fact, if you place a large number of edge devices and perform local processing, it performs better in real-time processing situations than in the cloud. On-device AI computing is advantageous in that it can efficiently process large-scale data through node distribution even when the data size is large. In addition, the advantages of on-device AI-based intelligent edge computing stand out even in environments where network connectivity is limited or unstable. Compared to cloud computing and edge computing, on-device AI computing does not rely heavily on network state as it handles the data of the device itself, providing a huge advantage, especially in environments where real-time processing is critical.

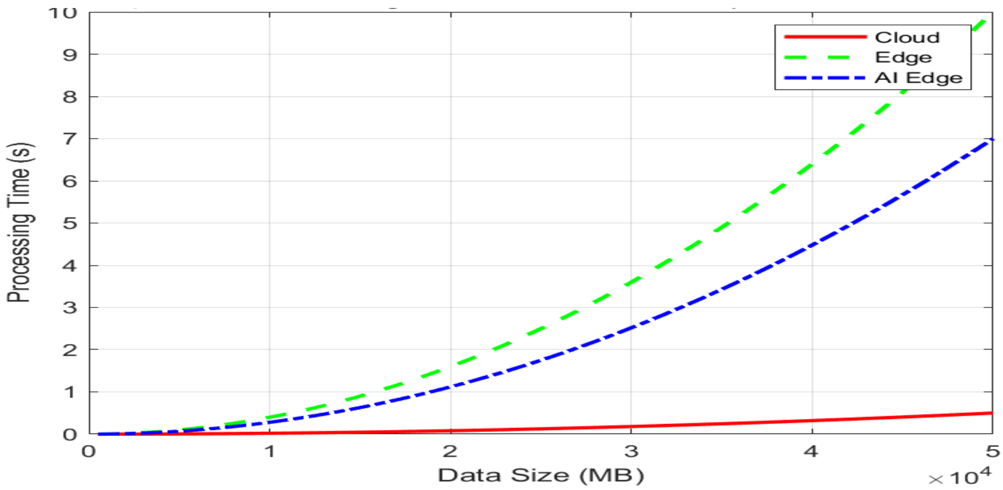

Figure 4. Average processing time by data size

3.3 Compute Processing Performance Comparison with Penalty Added

Figure 5 shows locally appearing processing time on the graph by applying additional infrastructure benefits and penalties such as distance, bottleneck, and node distribution where real data is processed. Edge computing and on-device AI-based intelligent edge computing have been shown to have relatively low actual latency compared to cloud computing due to local processing. On-device AI-based intelligent edge computing uses an intelligent ai system to reduce the amount of data transferred to minimize the latency seen locally. Cloud computing tends to experience a sharp increase in latency due to bottlenecks occurring in the process of processing all data and transferring it to a central server when penalty elements are added.

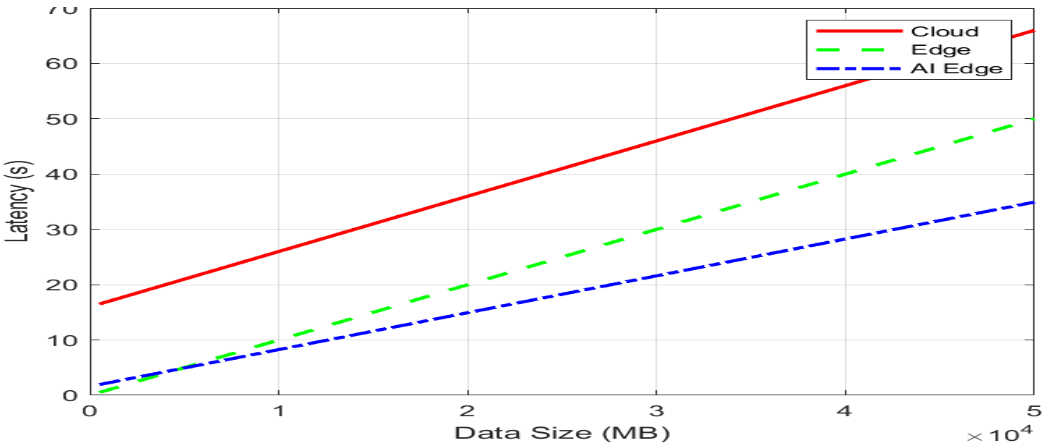

Figure 5. Average latency by Performance comparison with penalty added

Figure 5 shows the processing time that appears locally, taking into account additional infrastructure advantages and penalties such as the distance at which real data is processed, bottlenecks, and the distribution of nodes. In the case of cloud computing, the addition of these penalty elements tends to lead to a sharp increase in latency due to bottlenecks arising from the processing and transferring all data to a central server. On the other hand, edge computing and on-device AI-based intelligent edge computing have been shown to have relatively low latency despite infrastructure constraints due to the distributed structure. On-device AI-based intelligent edge computing processes data locally, reducing the actual amount of transmission data, minimizing processing time and preventing network bottlenecks through the distribution of nodes.

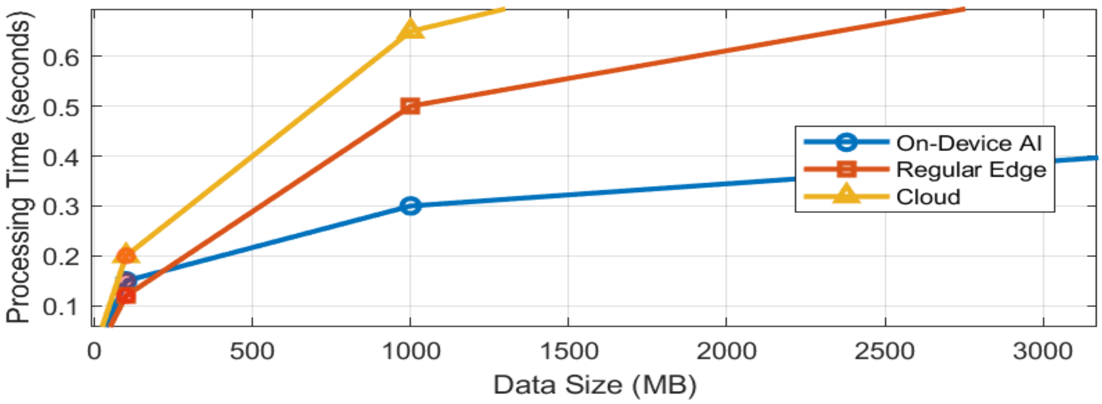

Figure 6. Average processing time by Performance comparison with penalty added

4. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

To summarize the experimental results, cloud computing, edge computing, and on-device AI-based intelligent edge computing have their own characteristics and advantages and disadvantages, which can be appropriately utilized for different situations and requirements. Cloud computing is optimized for strong processing performance and large-scale data processing, but requires high latency and reliable and fast network connectivity from network transmission. It is suitable for large-scale data processing, such as batch processing, but can have limitations in applications where real-time response is critical. Edge computing is advantageous for real-time data processing with low latency as it processes data at edge nodes close to the point where it occurs. It offers important advantages, especially in the field of IoT and real-time monitoring. However, it has limited computational performance compared to the cloud, which can lead to performance degradation in large-scale data processing, and stable network connectivity and security issues are important. Intelligent edge computing based on on-device AI shows particularly strong advantages. It is hardly affected by network bottlenecks or latency as it minimizes data transfer and performs real-time processing on local devices. In particular, it utilizes data compression technology to maximize data processing efficiency by reducing the amount of data to be transferred. These characteristics make it very advantageous in emergency situations where real-time processing is critical or in environments with unstable network connectivity. In addition, the security can be improved by processing data locally. However, even on-device AI-based intelligent edge computing has limitations in computational resources. Since the GPU and CPU performance of edge devices are relatively low compared to the cloud, performance degradation can occur during large-scale data processing. However, these limitations can be overcome through the use of data compression technology or the development of various on-device AI technologies.In conclusion, cloud computing, edge computing, and on-device AI-based intelligent edge computing can perform optimally in different environments and requirements. Cloud is ideal for large-scale data batch processing, and edge computing is suitable for situations where real-time data processing and low latency are critical.

On-device AI-based edge computing has advantages in applications where real-time processing performance and reduced network dependence are critical, and provides reliable performance even when the network is unstable.In the future, continuous research and development is needed to optimize the performance of these technologies and overcome their limitations, and advances in hardware and software technologies will enable more efficient and flexible data processing methods.

ACKNOWLEDGEMENTS

Research was conducted through the Hanseo University on-campus research support project.

REFERENCES

[1] Fernandes, V.; Carvalho, G.; Pereira, V.; Bernardino, J. Analyzing Data Reduction Techniques: An Experimental Perspective. Appl. Sci. 2024, 14, 3436.

[2] Siddiqui, S.T.; Khan, M.R.; Khan, Z.; Rana, N.; Khan, H.; Alam, M.I. Significance of Internet-of-Things Edge and Fog Computingin Education Sector. In Proceedings of the 2023 International Conference on Smart Computing and Application (ICSCA), Hail,Saudi Arabia, 5–6 February 2023; IEEE: Piscataway, NJ, USA; pp. 1–6

[3] Nasir Abbas at al "Mobile Edge Computing: A Survey" IEEE internet of Things Journal, Vol. 5, No. 1, pp. 450-465, Sep 2017.

[4] Y. Mao et al., "A Survey on Mobile Edge Computing: The Communication Perspective," IEEE Commun. Surveys Tuts, Vol. 19, no. 4, pp. 2322–2358, Apr. 2017.

[5] [J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," Future Generation Computer Systems, Vol. 29, No. 7, pp.1645-1660, Sept. 2013.

[6] Chen, A.; Liu, F.H.; Wang, S.D.e. Data reduction for real-time bridge vibration data on edge. In Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics, DSAA, Washington, DC, USA, 5–8 October 2019; pp. 602–603.

[7] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," Future Generation Computer Systems, Vol. 29, No. 7, pp.1645-1660, Sept. 2013.

[8] Kyung-rae. C ; Seok-min.H; Won-hyuk.C . Performance Comparison and Optimal Selection of Computing Techniques for Corridor Surveillance Journal of the Korean Society of Navigation, 2023, 27.6: 771-776.

[9] IFTIKHAR, Sundas, et al. AI-based fog and edge computing: A systematic review, taxonomy and future directions. Internet of Things, 2023, 21: 100674.

[10] HUA, Haochen, et al. Edge computing with artificial intelligence: A machine learning perspective. ACM Computing Surveys, 2023, 55.9: 1-35..

[11] JMERENDA, Massimo; PORCARO, Carlo; IERO, Demetrio. Edge machine learning for ai-enabled iot devices: A review. Sensors, 2020, 20.9: 2533.