AI-Based Fog and Edge Computing: A Comprehensive Systematic Review

Kaustubh Raykar
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra, Pune
raykarkaustubh@gmail.com

Hemakshi Bhatnagar
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra, Pune
bhatnagarhemakshi@gmail.com

Aamya Bansal
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra, Pune
aamyaa@gmail.com

Abhishek B
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra,Pune
abhishekbadapati@gmail.com

Mangal Singh
Dept of Electronics and Telecommunication Engineering
Symbiosis Institute of Technology
Maharashtra, Pune
mangal.singh@sitpune.edu.in

Prachi Nitin Kadam
Dept of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra, Pune
prachi.kadam@sitpune.edu.in

***Abstract*—This review paper aims to amalgamate artificial intelligence (AI) with fog and edge computing. This study offers insights into how these two technologies interact via real-time data and data analytics at the network edge. The study discusses the elements, applications, upcoming trends, and fog and edge computing shortcomings. Additionally, we are identifying potential directions and topics for further study. This paper is a significant resource for researchers, practitioners, and policymakers interested in using artificial intelligence and fog and edge computing.**

Keywords: Artificial Intelligence (AI); Fog Computing; Edge Computing; Real-time Data Analysis; Decentralized Computing; Cloud Resources; IoT (Internet of Things)

I. Introduction

The fusion of fog and edge computing with artificial intelligence marks a turning point in modern computing. The primary goal of this combination is to make it possible to allow real-time data analysis using the data sources. Fog and Edge computing components, such as fog nodes, cloud resources, and edge devices, serve as the infrastructure for enhanced computing in decentralized environments. Leading urban planning, healthcare, manufacturing, autonomous transportation, and agriculture applications have benefited. This systematic review seeks to summarize the state, achievements, difficulties, and potential directions for future research in the Fog and edge computing field.

II. Literature Review

Sundas Iftikhar, Sukhpal Singh Gill, Chenghao Song, and Minxian Xu (2022) conducted a systematic literature review of AI/ML approaches for fog/edge computing. They proposed a taxonomy of common AI algorithms used for resource management and compared existing studies based on infrastructure, objectives, platform, resource management, metrics, AI methods, and applications. Key findings were that evolutionary algorithms, machine learning, and combinatorial algorithms have been applied for fog application placement, with a focus on load balancing. Open issues identified include security considerations and the need for standardization. They had also various systematic reviews of papers and their findings presented important insights as shown in Fig. 1 [1]

Zahra Makki Nayeri (2021) reviewed application placement in fog computing using AI methods. A taxonomy was developed categorizing evolutionary algorithms, machine learning algorithms, and combinatorial algorithms. The advantages and disadvantages of different AI-based methods were discussed. It was found that evolutionary algorithms are commonly applied, focusing on hybrid meta-heuristic approaches. Machine learning, especially reinforcement learning and deep learning, optimizes performance metrics. Combining techniques can enhance solutions. Security and open challenges around scalability and reliability were noted.In order to provide a visual representation of the Fog Computing architecture, it is presented the following diagram (see Figure 2), which illustrates the three key layers of this hierarchical system: the terminal layer, the fog layer, and the cloud layer. [2] as shown in Fig 2 . **Fog Computing Architecture**

Year		2020	2019	2020	2022	2020	2019	2018	2019	2021	2022	2021	2022	2022
Environment	IoT		✓				✓	✓	✓					✓
	Edge			✓					✓					✓
	Fog	✓		✓		✓	✓	✓		✓	✓	✓	✓	✓
Empty Cell	Cloud		✓		✓				✓					✓
AI Method	Machine learning				✓				✓	✓				✓
	Deep learning									✓		✓		✓
	Reinforcement learning									✓	✓		✓	✓
AI for Fog/Edge	Resource Discovery												✓	✓
	Resource Estimation	✓											✓	✓
	Application Placement	✓				✓				✓			✓	✓
	Resource Orchestration		✓		✓								✓	✓
	Resource Scheduling	✓											✓	✓
	Resource Provisioning	✓			✓		✓		✓				✓	✓
	Task offloading	✓		✓									✓	✓
	Load balancing	✓											✓	✓
Taxonomy		✓	✓	✓	✓				✓		✓			✓
Classification					✓			✓	✓	✓		✓		✓
Systematic Literature Review (SLR)														✓

Fig. 1. Systematic Literature Review (SLR)

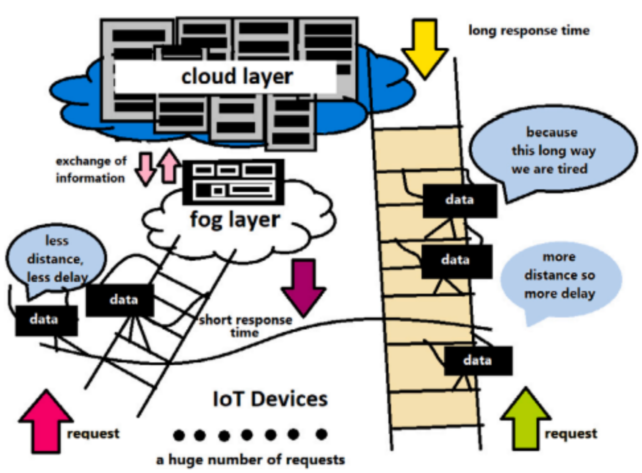

Fig. 2. Fog computing structure.

- **Terminal Layer:** The lowest layer, comprising IoT devices generating diverse data. These devices may be widely distributed and possess mobility features. Fog nodes in this layer act as sensors or have computational capabilities. Applications run here for near real-time responses.
- **Fog Layer:** The intermediate layer that connects IoT devices with the cloud. It extends closer to IoT devices for efficient interaction. Fog nodes in this layer support IoT devices and play a crucial role in executing real-time and latency-sensitive applications.
- **Cloud Layer:** The top layer with high processing and storage capabilities. It is centralized and associated with the fog layer. If the fog layer can't handle a request, it's forwarded to the cloud layer for more complex processing.

Comparison with Edge Computing:

- Edge computing is located on end devices, whereas Fog computing is closer to end devices like network switches and routers.
- Fog computing provides more resources, supports multiple IoT applications, and focuses on infrastructure rather than end devices.

IoT Applications: IoT applications involve common activities such as receiving data from IoT devices, preprocessing and analyzing received data, and handling events. Each application module performs specific operations and depends on resources like CPU, memory, and bandwidth for efficient data processing.Edge Computing's general architecture is of three layers which are end, edge, and cloud

Hong Wen (2020) provided a comprehensive overview of AI-based fog and edge computing, covering recent progress, methodologies, resource management, architectures, and applications. Key findings were that communication-efficient

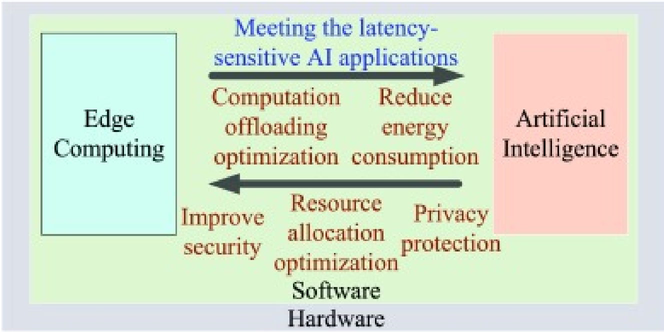

Fig. 3. Edge computing and AI

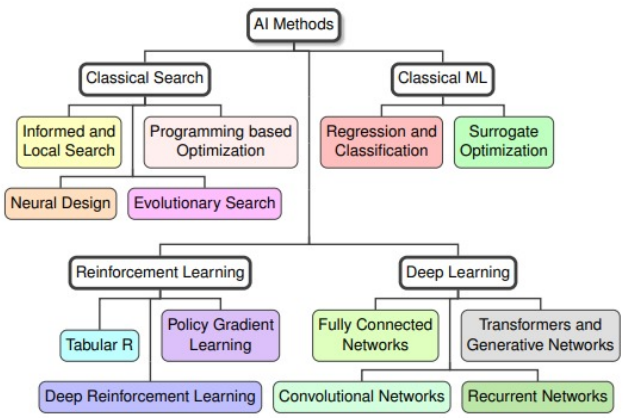

Fig. 4. Classification of Fog systems

edge training, optimal resource allocation, end-to-end architectures, and standardization are critical areas needing further research. [3]

Haochen Hua et al. (2023) surveyed the literature on edge computing and AI. They found that edge computing addresses limitations of cloud computing, and integrating AI/ML improves system performance and decision-making. Combining edge computing and AI/ML is a promising research direction.
In Figure 4 Edge computing and AI have a mutually beneficial interaction. The arrow pointing from right to left indicates that techniques, like those for computational offloading optimisation, are necessary to promote the advancement and improvement of edge computing. On the other hand, the left-to-right arrow suggests that edge computing should be located closer to end devices in order to meet the requirements of AI applications that are latency-sensitive, such as those found in smart city solutions. [4]
Shreshth Tuli et al. (2023) reviewed AI-augmented resource management in edge and fog computing, categorizing approaches into deployment, scheduling and maintenance strategies. Important Figures include Fig 4 a condensed classification of artificial intelligence techniques for fog systems that goes beyond Russell and Norvig's. Key observations were a shift towards deep learning models, digital twins, transformers, and unsupervised techniques. AI focuses on efficient deployment of models on edge devices. AI also facilitates resource management decisions like workload placement and scaling. Applications in healthcare, networking, manufacturing, and cities are emerging. Limitations around scalability, reliability, adaptability were noted. Future work should consider self-supervised models, analog AI, decentralized modeling, simulations, and co-design. [5]

Massimo Merenda et al. (2020) provided a literature review of edge computing and AI, discussing concepts, benefits, and challenges. Implementing AI technologies like machine learning and natural language processing in edge computing can enable automated computation and smart decisions. Key issues noted were the need for standardization, security, and privacy. [6]

Xiaoxu Ren et al. (2022) proposed a computing-power trading framework using blockchain to share resources between cloud/edge providers and AI consumers/miners. A Stackelberg game model balances profits. Simulations demonstrated resource utilization and profit balancing improvements. This enables flexible computing sharing and efficient AI service management. [7]

Luyi Chang et al. (2022) discussed integrating 6G edge intelligence into the Metaverse, investigating systems, core challenges like latency and security, and introducing advanced methods like federated learning to address them. Open research topics identified include cognitive workload and enhancing interoperability, security, and privacy. [8]

Yong Deng et al. (2020) provided a comprehensive overview of edge intelligence, analyzing AI for edge focusing on optimizing edge computing, and AI on edge studying training/inference on edge devices. A research roadmap was presented. Challenges include efficiency, coordination, data availability, and model selection. AI technologies can enhance efficiency. Applications include vehicles, smart homes/cities, and security. [9]

E. Badidi et al. (2023) reviewed edge AI and video analytics in smart cities, discussing applications, opportunities and challenges. Models, techniques, and methods were presented as in Fig 5 .Key issues noted were realizing the full potential of edge AI and video analytics. Distributing references by year and database was analyzed as well in Fig . 6 [10]

III. Methodology

In this section, we detail the systematic methodology employed in our research to investigate the amalgamation of fog and edge computing. Our methodology encompassed the

Fig. 5. Edge AI Systems

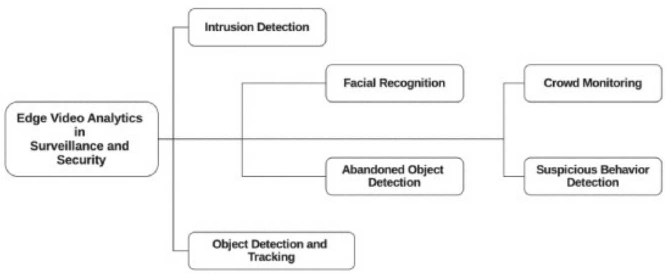

Fig. 6. Use cases of edge video analytics in surveillance and security.

following steps:

1. **Research Questions Formulation:** To frame our study, we formulated specific research questions that would guide our investigation into the interaction between fog computing, and edge computing. These questions played a crucial role in directing our search and analysis.

2. **Data Sources:** We conducted a rigorous search across a variety of reputable data sources to compile a comprehensive collection of relevant studies. Our search included databases such as Wiley Interscience, Springer, ACM Digital Library, IEEE Xplore, ScienceDirect, in addition to seeking supplementary resources like technical reports and edited books.

3. **Search Strings and Keywords:** We constructed search strings using keywords pertaining to fog/edge computing, challenges, metrics, simulators, and algorithms. These search strings were designed to maximize the identification of pertinent research publications and narrow down the available databases to the most relevant articles.

4. **Inclusion and Exclusion Criteria:** To ensure the selection of high-quality publications, we established a set of inclusion and exclusion criteria. This involved the exclusion of brief publications, non-peer-reviewed papers, book chapters, and low-quality studies. The rigorous application of these criteria helped maintain the integrity of our research.

5. **Data Extraction:** We extracted various data items from the selected articles, including paper identifier, publication date, bibliographic information, type of article, motivation, innovation, problem statement, resource management method, implementation environment, performance evaluation, workload type, performance metrics, and drawbacks. We also proactively reached out to authors for additional insights and information regarding their scholarly works.

IV. Areas of Future Exploration in Fog and Edge Computing

The fusion of fog and edge computing marks a significant convergence in modern computing. As we progress with our research in this dynamic field, we intend to explore several promising avenues to broaden the horizons of Fog and Edge Computing. The following are some areas of future exploration:

1. **Integration Across Diverse Domains:** Our research will delve deeper into the application of Fog and Edge Computing across various domains, such as urban planning, healthcare, manufacturing, autonomous transportation, and agriculture. We will address the unique challenges and sector-specific requirements to develop tailored solutions for real-world applications.

2. **Enhancing Real-time Data Analytics:** Building upon our emphasis on real-time data analysis at the network edge, we will explore advanced analytics techniques. This includes improving data processing, feature extraction, and optimizing the utilization of cloud resources and IoT (Internet of Things) devices for enhanced real-time decision-making.

3. **Efficiency in Decentralized Computing:** Future research will focus on optimizing resource provisioning, task offloading, and allocation techniques in Fog and Edge Computing environments. We will evaluate and enhance performance metrics to increase the overall efficiency.

4. **Quality of Service (QoS) Enhancement:** Improving QoS parameters, including latency, throughput, reliability, and energy efficiency, will remain a central aspect of our future research. We will develop intelligent algorithms and techniques to ensure efficient and reliable Fog and Edge Computing systems.

5. **Exploring Serverless Computing:** The integration of resource managers in serverless computing environments, particularly Function-as-a-Service (FaaS), presents exciting opportunities and challenges. We will explore cost-efficiency, resource optimization, and performance aspects in these environments.

6. **Promoting Environmental Sustainability:** Striking a balance between the resource requirements of edge devices

and resource-efficient computing is a significant concern. Future research will focus on developing energy-efficient models and algorithms to promote sustainability in Fog and Edge Computing.

7. **Advancements in Machine Learning Techniques:** We will explore and advance various machine learning techniques, such as Graph Neural Networks (GNN), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Recurrent Neural Networks (RNN), to enhance capabilities in decentralized computing environments.

8. **Online and Offline Machine Learning:** Investigating the use of online and offline machine learning approaches for Fog and Edge Computing will lead to improved resource management and decision-making. Our research will explore the benefits and limitations of these approaches and develop novel algorithms and frameworks.

V. Results and Discussions

In their 2020 research, Brogi and colleagues introduced a novel framework for addressing the application placement problem in Fog computing, focusing on two distinct perspectives: the algorithmic and modeling perspectives. The algorithmic perspective delves into methodological approaches and practical experiments to comprehensively understand the algorithmic aspects of this problem. Meanwhile, the modeling perspective emphasizes scrutinizing constraints and optimization metrics to identify suitable application placements.

Notably, their work stands out as it provides a thorough classification of AI techniques for Fog application placement, an aspect previously unexplored. Moreover, it offers a comprehensive evaluation of AI algorithms in this context, highlighting their strengths and weaknesses, bridging a significant gap in existing literature.

A. The Integration of AI and Edge Computing

The integration of AI and Edge Computing (EC) has two primary motivations: enhancing EC challenges by using AI to address issues like task scheduling, resource allocation, and privacy concerns, and optimizing AI applications by executing them on edge nodes for improved stability and user experience. The paper discusses the progress made in addressing these research issues, referencing relevant literature, and surveys in the field of Edge Computing and Artificial Intelligence.

B. Machine Learning Algorithms

Machine Learning Algorithms for Constrained Hardware

1. Deep Learning: Deep learning involves adjusting model parameters (weights and biases) using optimization functions, typically ADAM. The model is guided by an objective function (loss for supervised learning or reward for reinforcement learning) to identify patterns in data during training, making predictions. Different learning techniques include supervised, unsupervised, and reinforcement learning. Inference involves passing input data through network layers, with deep neural networks (DNNs) containing linear or nonlinear functions. Convolutional neural networks (CNNs) are used for image analysis, and recurrent neural networks (RNNs) handle sequential inputs.

2. Recurrent Neural Networks (RNNs): RNNs model dynamic temporal behavior, useful for tasks involving data sequences. Long short-term memory (LSTM) is a variant with memory control components for retaining values and information flow control.

3. Generative Adversarial Networks (GANs): GANs consist of a Generator and Discriminator to generate data and classify real data from synthetic data.

4. K-Nearest Neighbors Algorithm (K-NN): Used for pattern recognition, K-NN relies on nearby object features. ProtoNN is a modified version suitable for hardware-constrained devices.

5. Tree-Based ML Algorithms: Bonsai is an algorithm designed for resource-constrained IoT devices, maintaining prediction accuracy while minimizing model size and prediction costs.

6. Support Vector Machine (SVM): SVM is used for classification and regression, separating data classes with an optimal hyperplane, employing the kernel trick when data is not linearly separable.

C. Enabling Machine Learning at the Edge

1. Architectures: Three key architectures include on-device computation, edge server-based, and joint computation for low-latency model inference.

2. Model and Hardware Considerations: Model design focuses on fewer parameters for memory and latency reduction. Model compression techniques reduce computational demands. Hardware choice is influenced by factors like accuracy, energy efficiency, throughput, and cost, with micro controllers and compact hardware options suitable for IoT applications.

For a detailed list of hardware options for IoT devices implementing edge computing, refer to 7.

VI. Conclusion

In conclusion, this research paper, delves into the intersection of fog and edge computing, emphasizing key insights and areas of future exploration. The study brings forward several important findings and research directions:

1.Fog and Edge Computing Integration: The research paper offers a comprehensive review of the convergence of fog computing and edge computing. It highlights how these technologies interact, enabling real-time data analysis at the network edge. This convergence holds significant potential for various domains, including urban planning, healthcare, manufacturing, and more.

2.Machine Learning Algorithms for Resource-Constrained Hardware: The paper provides an overview of machine learning algorithms suitable for hardware with limitations. It discusses various algorithms that have applications in IoT devices and edge computing, highlighting their strengths.

Table 1. Hardware used for Internet of Things (IoT) devices that implement edge computing.

Work	DNN Model	Application	End Devices	Key Metrics
This work (Section 9)	CNN	Image Recognition	STM32F401RE (ARM® Cortex® -M4)	fast inference
[23]	SVM	Image Recognition	Raspberry Pi model 3 (ARM® v8)	fast inference
[90]	DNN	Distributed Computing	Raspberry Pi model 3 (ARM® v8)	hierarchical
[91]	SVM, CNN	Video Analysis	Raspberry Pi model 3 (ARM® v8)	fast inference
[92]	SVM	Video Analysis	Raspberry Pi model 3 (ARM® v8)	fast inference
[28]	SVM	Battery Lifetime Estimation	SPHERE	energy
[44]	CNN	Image Recognition, Sensor Fusion	Motorola 68HC11	fast inference
[65]	SVM	Code execution	ARM® v7	accuracy
[93,94]	Logistic Regression	Human Activity Recognition	ESP32	accuracy
[95]	CNN	Speech Recognition	Sparkfun Edge	accuracy

Fig. 7. Hardware for Iot Implementing Edge Computing

3.Enabling Machine Learning at the Edge: The research paper explores various architectures and considerations for enabling machine learning at the edge. It discusses on-device computation, edge server-based solutions, and joint computation to achieve low-latency model inference. It also emphasizes the importance of model design, model compression, and hardware choices for efficient edge computing.

4.Areas of Future Exploration: The paper identifies several promising avenues for future research in the field of fog and edge computing, including integration across diverse domains, enhancing real-time data analytics, improving efficiency in decentralized computing, enhancing quality of service (QoS), exploring serverless computing, promoting environmental sustainability, and advancements in computing techniques.

In summary, this research paper provides valuable insights into the integration of fog and edge computing. It highlights the potential for solving real-world challenges and suggests future research directions to advance this rapidly evolving field. Researchers, practitioners, and policymakers interested in fog computing and edge computing will find this paper to be a significant resource.

References

[1] Iftikhar, S., Gill, S. S., Song, C., Xu, M., Aslanpour, M. S., Toosi, A. N., Du, J., Wu, H., Ghosh, S., Chowdhury, D., Golec, M., Kumar, M., Abdelmoniem, A. M., Cuadrado, F., Varghese, B., Rana, O., Dustdar, S., & Uhlig, S. (2023). AI-based fog and edge computing: A systematic review, taxonomy and future directions. Internet of Things, 21, 100674. https://doi.org/10.1016/j.iot.2022.100674

[2] Nayeri, Z. M., Ghafarian, T., & Javadi, B. (2021). Application placement in Fog computing with AI approach: Taxonomy and a state of the art survey. Journal of Network and Computer Applications, 185, 103078. https://doi.org/10.1016/j.jnca.2021.103078

[3] Iftikhar, S., Gill, S. S., Song, C., Xu, M., Aslanpour, M. S., Toosi, A. N., Du, J., Wu, H., Ghosh, S., Chowdhury, D., Golec, M., Kumar, M., Abdelmoniem, A. M., Cuadrado, F., Varghese, B., Rana, O., Dustdar, S., & Uhlig, S. (2023). AI-based fog and edge computing: A systematic review, taxonomy and future directions. Internet of Things, 21, 100674. https://doi.org/10.1016/j.iot.2022.100674

[4] Haocen Hua , Yutong Li , Tonghe Wang , Nanqing Dong , Wei Li , Junwei Cao , Edge Computing with Artificial Intelligence: A Machine Learning Perspective (2023) . https://dl.acm.org/doi/10.1145/3555802

[5] Tuli, S., Mirhakimi, F., Pallewatta, S., Zawad, S., Casale, G., Javadi, B., Yan, F., Buyya, R., & Jennings, N. R. (2022). AI Augmented Edge and Fog Computing: Trends and Challenges. ArXiv. /abs/2208.00761

[6] Merenda, M., Porcaro, C., & Iero, D. (2019). Edge Machine Learning for AI-Enabled IoT Devices: A Review. Sensors, 20(9), 2533. https://doi.org/10.3390/s20092533

[7] X. Ren et al., "AI-Bazaar: A Cloud-Edge Computing Power Trading Framework for Ubiquitous AI Services," in IEEE Transactions on Cloud Computing, vol. 11, no. 3, pp. 2337-2348, 1 July-Sept. 2023, doi: 10.1109/TCC.2022.3201544.

[8] L. Chang et al., "6G-Enabled Edge AI for Metaverse: Challenges, Methods, and Future Research Directions," in Journal of Communications and Information Networks, vol. 7, no. 2, pp. 107-121, June 2022, doi: 10.23919/JCIN.2022.9815195.

[9] Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., & Zomaya, A. Y. (2019). Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. ArXiv. https://doi.org/10.1109/JIOT.2020.2984887

[10] E. Badidi, K. Moumane and F. E. Ghazi, "Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review," in IEEE Access, vol. 11, pp. 80543-80572, 2023, doi: 10.1109/ACCESS.2023.3300658.

[11] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang. 2019. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proc. IEEE 107, 8 (2019), 1738–1762.

[12] M. Miranda, C. Cristina, and S. Sebastiá. 2020. Deep learning at the mobile edge: Opportunities for 5G networks. Appl. Sci. 10, 14 (2020), 4735.

[13] F. Wang, M. Zhang, X. Wang, X. Ma , and J. Liu. 2020. Deep learning for edge computing applications: A state-of-theart survey. IEEE Access 8 (2020), 58322–58336.

[14] J. Chen and X. Ran. 2019. Deep learning with edge computing: A review. Proc. IEEE 107, 8 (2019), 1655–1674.

[15] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen. 2020. Convergence of edge computing and deep learning: A comprehensive survey. IEEE Commun. Surv. Tut. 22, 2 (2020), 869–904