Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/87/20241582

The practical applications of federated learning across various domains

Hanjing Wang

College of Information Science and Engineering, Ocean University of China- Ocean University of China, Qingdao, China

whj4421@stu.ouc.edu.cn

Abstract. With the advancement of artificial intelligence technology, a vast amount of data is transmitted during the model training process, significantly increasing the risk of data leakage. In an era where data privacy is highly valued, protecting data from leakage has become an urgent issue. Federated Learning (FL) has thus been proposed and applied across various fields. This paper presents the applications of FL in five key areas: healthcare, urban transportation, computer vision, Industrial Internet of Things (IIoT), and 5G networks. This paper discusses the feasibility of implementing FL for privacy protection in the aforementioned five real-world application scenarios and analyzes its accuracy and effiency. Additionally, it compares the FL framework with traditional frameworks, exploring the improvements FL has made in terms of privacy protection and performance, as well as the existing shortcomings of the FL framework. Further discussions are provided on potential future improvements. Moreover, this paper offers an outlook on current research trends and the developmental prospects in this research field.

Keywords: Federated learning, privacy-preserving, efficiency.

1. Introduction

Privacy issues are one of the major concerns today. The development of big data, artificial intelligence, and other technologies has inevitably led to problems related to data privacy breaches. User data is transmitted during the use of various software, websites, etc.; if this information is leaked, it can lead to various illegal activities such as fraud and extortion. The ability of enterprises to protect user privacy data significantly affects users' trust in them. With technological advancement, an increasing number of devices are being utilized. Currently, nearly 7 billion connected devices are used in Internet of Things (IoT) [1], and the number of smartphone users has almost reached 3 billion [1]. Consequently, the volume of data transmission between devices has greatly increased. In the field of deep learning, a substantial amount of data is collected and utilized for training deep models. While computational power and time have garnered widespread attention, data privacy issues were initially overlooked. The increase in data transmission volume can easily lead to serious privacy breaches [2]. In certain fields, the transmitted data often involves industry secrets and user privacy; if such data is leaked, it could lead to severe incidents, thus drawing significant attention to data security issues.

To safeguard data privacy, researchers have undertaken various attempts, among which the introduction of the Federated Learning (FL) model stands out as a significant approach. Researchers are not only delving deeply into FL model algorithms but also exploring potential real-world applications

DOI: 10.54254/2755-2721/87/20241582

of FL across multiple domains. A critical question they address is how to utilize FL models to protect data privacy without compromising model accuracy. This paper aims to review the applications of FL in the fields of medicine, urban transportation, visual systems, Industrial Internet of Things (IIoT), and 5G networks. It analyzes and summarizes the performance of FL in data privacy protection. Additionally, the paper organizes and examines methods to enhance FL's performance in terms of accuracy and data processing efficiency. Furthermore, it provides an outlook on the future prospects of FL applications.

2. The Theory of FL

To overcome the limitations posed by data privacy on artificial intelligence, FL was proposed to safeguard user privacy [3]. Research on FL is still in its infancy, and many scholars are conducting studies in this field. Overall, FL is an iterative process [4], in which data are brought into the code. FL is implemented through three steps: (1) initiate a global model [4]; (2) training the initial machine learning (ML) models on clients using personal data [4]; (3) training the local models at the client level, updating them, and sending the updates to the server where they are aggregated and used in order to update the global model [4]. After that, the newly updated model is transmitted back to each client [4]. Steps (2) and (3) are repeated [4]. Based on the aforementioned steps, there is no data transmission between clients. The data of each client is kept locally, which helps to protect the privacy of user data. The above steps are depicted in Figure 1.

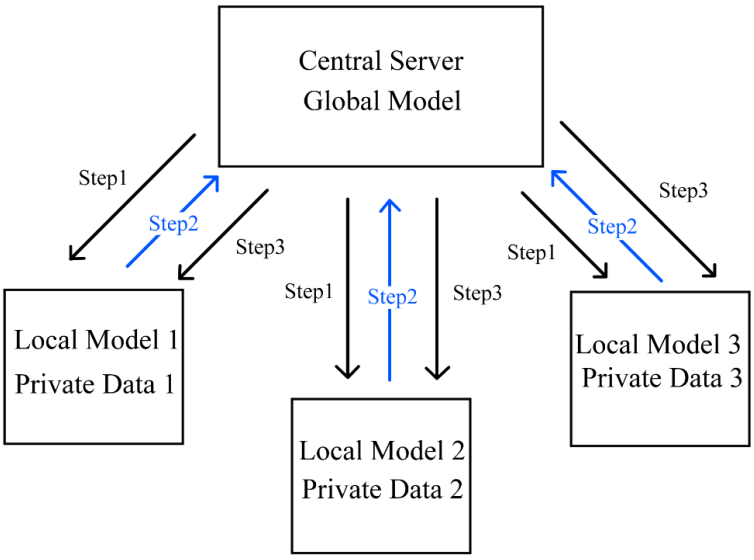

Figure 1. The three steps of FL(Picture credit : Original)

FL can be categorized in various ways, including network topology, data partitioning, open-source frameworks, data availability, and optimal aggregation algorithms [4]. Based on network topology, common classifications of FL are Centralized & Clustered FL and Fully-Decentralized FL [4]. Centralized & Clustered FL relies on a single central server, while Fully-Decentralized FL uses multiple coordinating nodes and clusters for distributed aggregation. According to data partitioning, it can be split into the following three types: Vertical FL, Horizontal FL, and Transfer FL [4]. For instance, SecureBoost, which combines XGBoost and FL, falls under this category. Figure 2 illustrates the specific categorization methods of FL [4].

DOI: 10.54254/2755-2721/87/20241582

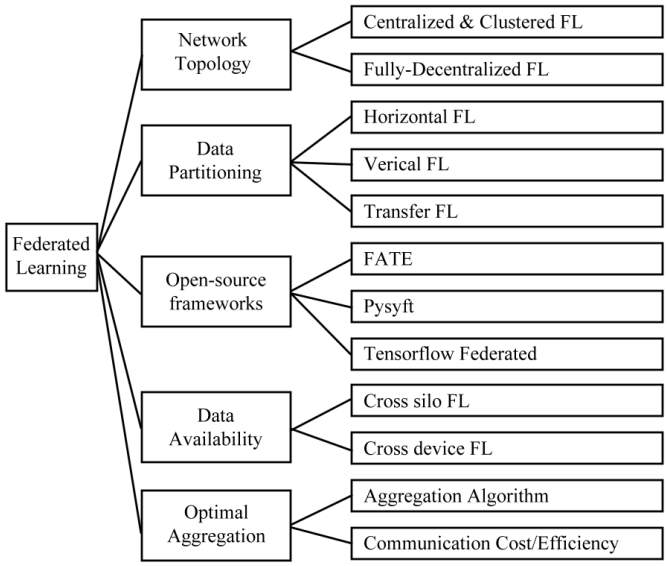

Figure 2. The Classification of FL[4]

FL data often exhibits authenticity and privacy, with labels that can be directly obtained [3]. In the real world, most domains involving private data receive considerable attention, and many scenarios involve high data transmission costs or require distributed collaborative training (e.g., intelligent transportation and autonomous driving). Data in the fields of medicine, urban transportation, vision, IIoT, and 5G networks often align with the characteristics required for FL data. Numerous scholars have conducted application research on FL in these domains.

3. Application Analysis

In various fields, data privacy and security hold paramount importance. This paper will discuss and summarize the applications of FL in five key areas: healthcare, urban transportation, computer vision, IIoT, and 5G networks, and analyze its performance in these domains.

3.1. Medicine

In the medical field, data often possesses a high degree of privacy and sensitivity [5]. This data includes patients' personal identification information and health information, making the confidentiality of medical data particularly crucial.

To protect user privacy, reference [5] utilized the characteristic of FL where data does not need to be uploaded to a central server to train the learning model, while reference [6] employed FL to train local datasets from different sites. The datasets used in references [7, 8] consisted of 120 samples with six different attributes (urinary pain, urethral burning sensation, itching and swelling, urgency, occurrence of nausea, discomfort in the lower back, and body temperature). Reference [6] compared traditional ML methods with FL g methods. To ensure model accuracy while safeguarding data privacy, researchers used datasets in text file format and preprocessed the data. Experimental data from reference [6] indicates that compared to traditional ML approaches, FL not only enhances data privacy but also achieves nearly 100% accuracy.

The researchers in reference [9] utilized a multimodal FL model to evaluate the diagnostic efficacy in gynecologic malignancies, encompassing a dataset of over 500 patients. Figure3 illustrates the process of multimodal FL [10]. In addition to employing the FL model, reference [9] implemented stringent access controls, restricting access to authorized researchers and medical personnel only, who could access patient privacy data under authorization, and anonymized patient data (such as removing identifiable information that could disclose patient identities). Through these methods, the researchers enhanced data privacy protection, offering a feasible approach. Reference [9] partitioned the dataset into

DOI: 10.54254/2755-2721/87/20241582

two parts (training data sets and testing data sets) to leverage the advantages of FL, demonstrating that FL's capability to effectively safeguard data privacy played a crucial role. Table 1 illustrates the performance comparison between traditional methods and multimodal FL [9]. The comparison reveals that multimodal FL enhances sensitivity while safeguarding patient privacy, underscoring its significant potential as a more effective diagnostic tool in the field of gynecologic malignancies.

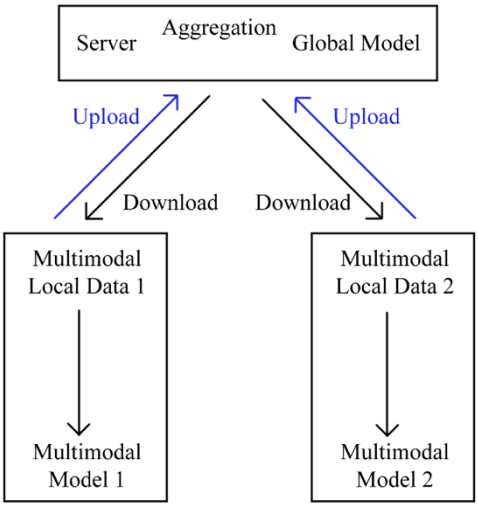

Figure 3. The process of Multimodal FL[10]

Table 1. Traditional method v/s Multimodal FL

	Traditional (CT)	Traditional (MRI)	Multimodal FL Without Image Information	Multimodal FL With Image Information (MRI)
Sensitivity	0.3263	0.359	0.923	0.941
Specificity	0.9215	0.9337	0.922	0.967

3.2. Urban Transportation

In the domain of urban transportation, directly collecting user information may expose their privacy, such as personal identification details, geographical locations, and mobility patterns. The protection of data privacy is directly correlated with the trustworthiness of users.

To guarantee individual data security, reference [11] proposed the DRLE framework to establish a decentralized learning edge computing approach, yet there still exists certain risks during the collection of raw vehicle data [12]. In order to augment the security of data privacy, reference [12] considered employing FL for model training. Meanwhile, to analyze the feasibility of this approach, researchers utilized the same model as reference [13] as a benchmark for comparison. The final results obtained are depicted in Table 2 [12,13]. From the comparison, it is evident that while the privacy of the data has been improved, the accuracy of the proposed model decreased from 76.81% to 71.02%.

Table 2. Baseline v/s FL

	Baseline	FL
Accuracy	0.7681	0.7102
Precesion	0.7681	0.7002
Recall	0.7681	0.7085
F1-Score	0.768	0.7100

DOI: 10.54254/2755-2721/87/20241582

FL's integration and Transport Mode Inference (TMI) was proposed by reference [14] to enhance data privacy, termed as PPDF-FedTMI. To assess the model's performance, researchers utilized a GPS-based dataset [15] and reconstructed its trajectories in experimental preparation. Analysis of the experimental results [14] regarding metrics demonstrating significant potential. However, there is still a need for further development in balancing privacy and utility aspects [14].

3.3. Visual System

In the field of computer vision, data privacy is of paramount importance. For instance, data utilized in applications such as facial recognition and surveillance systems often involve sensitive information like users' personal identities. If this information is compromised, it could potentially lead to malicious events such as identity theft, resulting in economic and psychological losses for the users.

Reference [16] addresses the unique needs of individuals with hearing impairments by leveraging FL to detect Bengali Sign Language while preserving user privacy. Researchers in [16] established a FL framework and conducted a comprehensive evaluation of six models concerning accuracy, precision, F1 score, recall, and loss. Among these, the proposed Federated Averaging (FedAVG) model achieved an accuracy of 98.36% (correctly predicting 9246 out of 9400 samples) [16]. The implementation process of FedAVG is illustrated in Figure 4 [17]. In this experiment, FL demonstrated its capability to effectively protect data privacy while achieving high accuracy. However, there are still some privacy risks associated with collaborative model training that need to be addressed (such as The risks associated with collaborative training participants [16]).

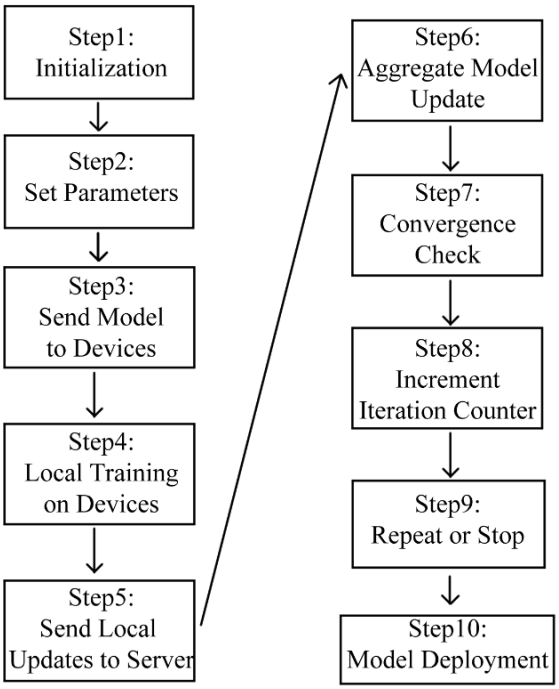

Figure 4. The implementation process of FedAVG [17]

In addition to gesture recognition, FL has also shown promise in human body posture recognition. Reference [18] proposed a FL framework (FL-HPR) for human posture recognition, aiming to protect data privacy. In the study conducted by reference [18], the researchers performed five-fold cross-validation on the client-side performance of three FL models (FedAVG, Fedprox, and FedBN) and found that the FL framework successfully optimized point cloud segmentation networks' average performance while protecting data privacy. In this study, human posture images were either unobstructed or

DOI: 10.54254/2755-2721/87/20241582

minimally obstructed. The researchers anticipate achieving high accuracy in recognition even under conditions of severe occlusion .

3.4. IIoT

In the IIoT domain, data privacy is equally paramount. Data encompassing sensitive business information such as operational efficiency, equipment parameters, and manufacturing processes is involved. Unauthorized access to this data could inflict significant economic losses on enterprises, and in malicious attack scenarios, compromise production processes leading to security incidents.

To mitigate potential data security issues, researchers consider adopting a decentralized architecture, namely the FL model. The study by researchers in reference [19] integrates FL methods to achieve large-scale distributed deep learning in IoT environments, ensuring user privacy protection and efficient communication. Reference [19] primarily employs approximate computing, distributed optimization algorithms, incremental learning, and differential privacy techniques, among others, to test three real-world datasets. Ultimately, reference [19] achieves a privacy preservation accuracy of 98%, marking an improvement over traditional privacy protection techniques, while also enhancing communication efficiency. However, its resilience against potential interference attacks requires further enhancement.

Regarding the aforementioned issues, reference [20] conducted a more in-depth investigation. To safeguard data privacy against threats such as data poisoning attacks and interference attacks, reference [20] researchers proposed a FL model using multiparty computation. FL faces various threats including interference attacks, data leakage, and model reverse engineering, as illustrated in Figure 5 [20]. Given these myriad threats, upgrading FL systems becomes imperative. Reference [20] employed algorithms based on secure multiparty computation to facilitate collaborative computation among multiple parties while ensuring their respective data's privacy. Compared to conventional FL algorithms, it achieved enhanced accuracy; however, an increase in the number of clients also led to higher communication overhead and latency rates compared to traditional FL algorithms.

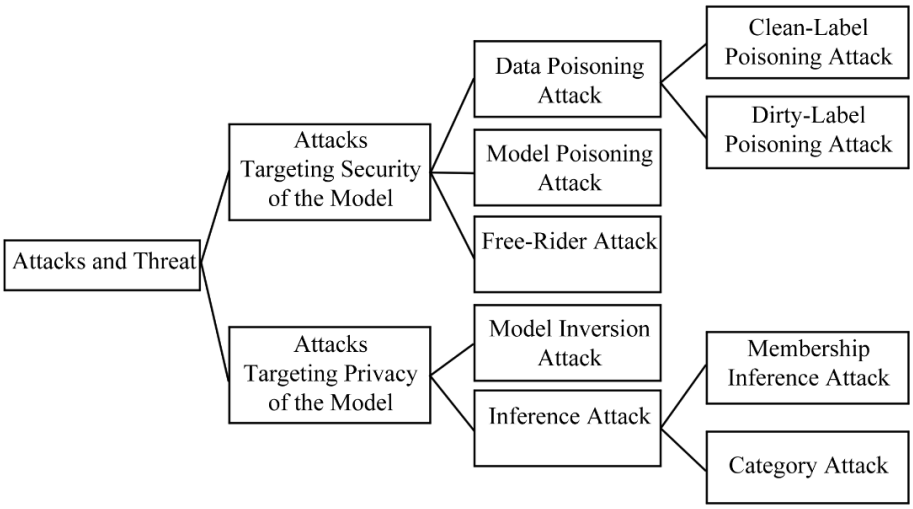

Figure 5. The classification of threats faced by FL[20]

3.5. G Networks

With the widespread adoption of 5G networks, a vast amount of information concerning user privacy and corporate confidentiality is transmitted through these networks. While 5G networks facilitate rapid data processing, they also escalate the risk of malicious attacks. Therefore, safeguarding the data transmitted over 5G networks is of paramount importance.

The researchers in reference [17] addressed privacy protection concerns by integrating FedAvg, adaptive learning rate, and secure aggregation for collaborative model training. Compared to traditional methods such as decision trees and linear regression, the approach proposed in reference [17]

DOI: 10.54254/2755-2721/87/20241582

demonstrates significantly superior accuracy (achieving 95.2%) while also preserving data privacy. Furthermore, it exhibits higher efficiency in data processing, meeting real-time requirements and utilizing memory resources more effectively than methods like logistic regression [17]. Table 3 illustrates the comparative performance of FL versus traditional methods [17]. In the realm of network applications, FL models exhibit exceptional adaptability and hold promising prospects for broader future applications.

Table 3. Proposed FL Method v/s Traditional Method

Method	Accuracy (%)	Efficiency (ms)	Memory Usage (MB)	Scalability (Nodes)
Proposed FL Method	95.2	25	120	5000
Linear Regression	88.5	30	----	----
Decision Trees	92.1	35	----	----
Logistic Regression	----	----	150	4000
Random Forest	----	----	200	3500

Introducing artificial intelligence algorithms in 5G networks raises significant concerns about privacy protection. To effectively safeguard user data privacy, researchers in [21] proposed an asynchronous weight updating framework based on FL. This framework is split into two distinct parts: client-side training and central node training, allowing clients to locally update their model parameters and optimizing them on network slicing [21]. As the number of clients increases, reference [21] demonstrated stable performance improvements, with the model's performance also increasing with additional time rounds. Reference [21] achieved a low-latency approach that reduces overhead while enhancing throughput, demonstrating high-performance applications in the 5G domain.

4. Conclusion

As a newly emerged technology, FL has successfully contributed to privacy protection. This paper provided a brief introduction to the principles of FL, shifting their focus from theoretical aspects to practical applications. The paper organized, analyzed, and summarized the applications of FL in five domains: medicine, urban transportation, visual systems, IIoT, and 5G networks. The findings revealed that FL can effectively protect data privacy, demonstrating superior performance in this regard. Accuracy, a crucial metric for FL models, can be ensured through the integration of other algorithms, model optimization, and data preprocessing. Compared to traditional deep learning models, FL models can achieve significant improvements in accuracy. Additionally, performance metrics such as memory consumption can be enhanced by optimizing certain aspects of FL, such as network slicing. In practical applications, optimized FL models outperform traditional deep learning models concerning privacy protection, data processing efficiency, and accuracy. The optimization direction of FL varies depending on the specific domain and requires contextual judgment. Overall, FL has broad application prospects in real life. Beyond specific domains, researchers can delve deeper into hybrid fields for further exploration.

References

[1] Lim, W. Y. B., et al. 2020 "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," IEEE Commun. Surv. Tutorials, vol. 22, no. 3, pp. 2031-2063.

[2] Fang, C., et al. 2022 “A privacy-preserving and verifiable federated learning method based on blockchain”, Comput. Commun., vol. 186, pp. 1-11.

[3] McMahan, H. B., et al. 2016 “Communication-Efficient Learning of Deep Networks from Decentralized Data”. arXiv preprint arXiv:1602.05629.

[4] Mothukuri, V., et al. 2021 “A survey on security and privacy of federated learning”, Future Gener. Comput. Syst., vol. 115, pp. 619-640.

[5] Shah, U., et al. 2021 "Maintaining Privacy in Medical Imaging with Federated Learning, Deep Learning, Differential Privacy, and Encrypted Computation," 2021 6th Int. Conf. for Convergence in Technol. (I2CT), Maharashtra, India, pp. 1-6.
[6] Shah, H., Patel, R., and Tawde, P. 2023 "Federated Learning to Preserve the Privacy of User Data," 2023 Somaiya Int. Conf. on Technol. and Inf. Manag. (SICTIM), Mumbai, India, pp. 23-27.
[7] Czerniak, J. and Zarzycki, H. 2003 "Application of rough sets in the presumptive diagnosis of urinary system diseases", Artif. Intell. and Security in Comput. Syst., ACS'2002 9th Int. Conf. Proc., Kluwer Acad. Publ., pp. 41-51.
[8] UCI Machine Learning Repository: –Acute Inflammations Data Set.
[9] Hu, Z., et al. 2023 "Comparison of Multi-Modal Federated Learning Framework and SPSS in the Evaluation of Lymph Node Metastasis Probability in Gynecological Malignancies," 2023 IEEE 4th Int. Conf. on Pattern Recognit. and Mach. Learn. (PRML), Urumqi, China, pp. 280-284.
[10] Lin, Yi-Ming, et al. 2023 Federated Learning on Multimodal Data: A Comprehensive Survey, MIR, 20(4): 539-553.
[11] Zhou, P., et al. 2021 DRLE: Decentralized reinforcement learning at the edge for traffic light control in the IoV, IEEE Trans. on Intell. Transp. Syst., vol. 22, pp. 2262–2273.
[12] Gomes, G. L., da Cunha, F. D., and Villas, L. A. 2023 "Differential Privacy: Exploring Federated Learning Privacy Issue to Improve Mobility Quality," IEEE Latin-Am. Conf. on Commun. (LATINCOM), Panama City, Panama, pp. 1-6.
[13] Wang, S. 2018 Traffic jam prediction hackntx. Accessed: 2023-09-29.
[14] Huang, Qihan, et al. 2023 PPDF-FedTMI, A Federated Learning-based Transport Mode Inference Model with Privacy-Preserving Data Fusion, Simul. Model. Pract. and Theory, vol. 129, Art. 102845.
[15] Zheng, Y.,W. Y. B. Lim et al., "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," in IEEE *Commun. Surv.* Tutorials, vol. 22, no. 3, pp. 2031-2063, thirdquarter 2020
[16] Sarkar Diba, Bidita, et al. 2024 Explainable federated learning for privacy-preserving bangla sign language detection, Engineering Applications of Artificial Intelligence, vol. 134, p. 108657.
[17] Ojha, A. C., Yadav, D. Kumar, and B, A. 2023 "Federated Learning Paradigms in Network Security for Distributed Systems," 2023 IEEE Int. Conf. on ICT in Bus. Ind. & Gov. (ICTBIG), Indore, India, pp. 1-5.
[18] Wang, Jiaxin, et al. (2024) Multi-sensor fusion federated learning method of human posture recognition for dual-arm nursing robots, Inf. Fusion, vol. 107, Art. No. 102320.
[19] Du, W., et al. "Approximate to Be Great: Communication Efficient and Privacy-Preserving Large-Scale Distributed Deep Learning in Internet of Things," IEEE Internet Things J., vol. 7, no. 12.
[20] Huang, R.-Y., Samaraweera, D., and Chang, J. M. 2023 "Exploring Threats, Defenses, and Privacy-Preserving Techniques in Federated Learning: A Survey," Computer, vol. 57, no. 4, pp. 46-56.
[21] Bedda, K., Fadlullah, Z. M., and Fouda, M. M. 2022 "Efficient Wireless Network Slicing in 5G Networks: An Asynchronous Federated Learning Approach," 2022 IEEE Int. Conf. on Internet of Things and Intelligence Systems (IoTaIS), BALI, Indonesia, pp. 285-289.