DOI. No. 10.13052/2794-7254.007

Federated Learning for Energy Efficiency in 6G

Satwat Bashir, Tasos Dagiuklas, Kasra Kassai and Muddesar Iqbal*

Abstract: This paper presents a multi-tier Federated Learning (FL) architecture designed to optimize energy efficiency in 6G, with particular emphasis on compliance with the Network Data Analytics Function (NWDAF) standards defined by 3GPP. Unlike existing FL architectures that often overlook energy efficiency and lack full integration with network functions like NWDAF, our proposed architecture integrates AI-driven strategies across multi layers. This multi-tier approach dynamically adjusts computation and communication rounds, reducing energy consumption while maintaining high model accuracy and network performance. By addressing challenges such as data heterogeneity and personalisation through adaptive training, intelligent routing, and advanced model aggregation, the architecture significantly enhances energy efficiency. Initial simulations, aligned with NWDAF processing requirements, underscore the architecture's suitability for deployment in 6G , offering a scalable, energy-efficient, and privacy-preserving solution that aligns with industry standards and addresses key challenges in distributed learning.

Keywords: 6G, energy efficiency, federated learning, multi-tier architecture, NWDAF, AI-driven strategies.

1. Introduction

The advent of 5G has led to a significant increase in data generation and the need for real-time analytics. In response, the Third Generation Partnership Project (3GPP) introduced the Network Data Analytics Function (NWDAF) in Release 17, 18 and 19 [1–3] to manage, optimize and automatise network operations through data-driven insights. NWDAF facilitates the collection and analysis of network data to enhance performance, reliability, and user experience. However, the distributed nature of data and the requirement for privacy-preserving mechanisms present substantial challenges that need to be addressed.

As we move towards 6G, the focus extends beyond enhancements in data throughput and connectivity. A critical aspect of 6G development is the integration of advanced intelligence to ensure sustainable and efficient network operations [4–6]. Energy efficiency and sustainability have emerged as paramount concerns, particularly as networks become increasingly complex and data-intensive [7]. Consequently, optimizing energy consumption in communication networks is a central priority in 6G technologies.

Edge computing has emerged as an effective strategy to address these energy challenges. By utilizing a combination of platforms, Internet of Things (IoT) devices, and software, edge computing systems efficiently manage computational resources, distributing processing tasks according to workload requirements. By offloading computational demand from centralized servers, edge architectures conserve bandwidth, alleviate cloud infrastructure load, and reduce latency. This approach is critical for real-time applications like autonomous vehicles and smart cities, where immediate processing is essential.

Despite its advantages, edge computing introduces challenges such as increased security risks due to its distributed nature, bandwidth management complexities, and hardware heterogeneity. Managing these issues is essential to ensure consistent performance and secure data processing across devices.

Federated Learning (FL) has emerged as a promising approach to address some of these challenges by enabling local model training on devices and aggregating updates centrally. This method preserves user privacy and reduces the need to transmit sensitive raw data, aligning with the privacy requirements of modern networks. However, FL faces issues like communication overhead from frequent model updates, dependence on a central server for aggregation, and performance degradation when client devices hold non-independent and identically distributed (non-IID) data.

Combining edge computing with FL creates a synergistic relationship that enhances the capabilities of both technologies. Edge computing provides the necessary infrastructure for FL to operate more efficiently, reducing latency and communication costs. This synergy leads to a more robust and scalable system for managing distributed data.

Several FL architectures have been proposed, leveraging the benefits of edge computing to improve privacy preservation, reduce communication overhead, and enhance scalability [8, 9]. However, these architectures often overlook energy efficiency and do not fully integrate with network functions like NWDAF [10, 11]. Additionally, challenges related to data heterogeneity and personalisation remain inadequately addressed [9].

To bridge these gaps, we evaluate how a previously proposed multi-tier FL architecture [12], designed to handle data heterogeneity and personalization, fits within 6G while aligning with the NWDAF framework defined by 3GPP standards [3]. By integrating AI-driven strategies across its Client, Edge, Fog, and Cloud layers, the architecture dynamically adjusts computational resources to reduce energy consumption while maintaining

Department of Computer Science and Informatics
London South Bank University, London, UK
E-mail: bashis11@lsbu.ac.uk; tdagiuklas@lsbu.ac.uk; kasra.kassai@lsbu.ac.uk; m.iqbal@lsbu.ac.uk
*Corresponding Author
Manuscript received 03 October 2024, accepted 25 October 2024, and ready for publication 21 December 2024.

model accuracy and network performance. This design effectively manages data heterogeneity and personalisation while minimizing redundant computations and communication overhead.

In this paper, we evaluate the energy efficiency of the proposed multi-tier FL architecture, particularly highlighting how it seamlessly integrates with NWDAF in accordance with 3GPP standards, making it a well-aligned solution for optimizing 6G. We demonstrate how adaptive training, intelligent routing, and advanced model aggregation contribute to significant energy savings. Initial simulations highlight the effectiveness of these strategies in 6G, suggesting substantial improvements in energy efficiency for real-world applications.

The remainder of this paper is organized as follows: Section II details the proposed multi-tier FL architecture, discussing the specific roles and contributions of each layer in enhancing energy efficiency and aligning with NWDAF. Section III presents the results of our simulations, demonstrating the framework's effectiveness in reducing energy consumption while maintaining model accuracy. Section IV discusses the AI-driven strategies employed across the layers to optimize energy efficiency. Finally, Section V concludes the paper and outlines potential areas for future research and real-world implementation in diverse 6G scenarios.

2. Multi-Tier Federated Learning Architecture and its Integration with NWDAF

The multi-tier FL architecture, originally proposed in [12], comprises four hierarchical layers: **Client**, **Edge**, **Fedge**, and **Cloud** as shown in Figure 1. This architecture is designed to manage data and model training effectively in distributed networks, addressing challenges related to data heterogeneity, personalisation, and energy efficiency in 6G environments. In this section, we review this architecture from the perspective of the NWDAF, highlighting how it aligns with 3GPP standards and can be seamlessly integrated into existing network infrastructures.

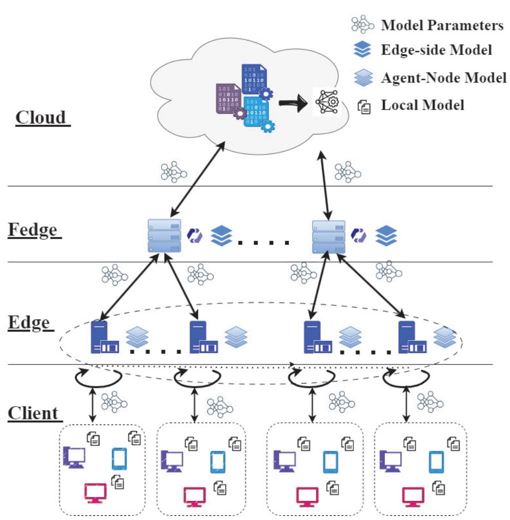

Figure 1.
Proposed multi-tier federated learning architecture.

2.1. Overview of the Architecture

The proposed architecture combines the capabilities of edge computing and FL to distribute computational tasks and data processing efficiently across different layers of the network. This approach reduces communication overhead and redundant computations, improving energy efficiency and performance in model training. NWDAF plays a crucial role in this architecture by enabling distributed data analytics and optimizing network operations, ensuring privacy and network efficiency.

In this architecture, NWDAF facilitates both vertical (North/South bound communication) and horizontal (East/West bound communication) data flow to support real-time analytics across various network functions such as Application Function (AF), Session Management Function (SMF), Access and Mobility Management Function (AMF), (Network Exposure Function) NEF, and Policy Control Function (PCF), as depicted in Figure 2. The Hyperscaler provides additional computational power, allowing NWDAF to leverage machine learning (ML) models embedded across the network for optimizing resource allocation and reducing latency. This interaction ensures that the FL system operates efficiently in compliance with 3GPP standards and effectively adjusts to varying network conditions.

2.2. Client layer

The Client layer consists of end-user devices, such as smartphones and IoT devices, where local model training occurs using the client's own data. This allows for privacy-preserving operations, as raw data remains on the device, which aligns with 3GPP privacy requirements.

Regarding NWDAF integration, clients primarily interact with the network through standardized interfaces. These devices share aggregated model updates, not raw data, with the higher network layers, particularly the Edge layer, to support distributed analytics. In line with NWDAF's objective of safeguarding privacy, these communications are carried out securely, ensuring that the data is handled in compliance with privacy regulations.

Energy efficiency is enhanced by minimizing data transmission to the Edge Layer. Clients send only critical model updates, which reduces redundant communications. Additionally, clients utilize adaptive training techniques to optimize learning and minimize resource consumption. By dynamically adjusting learning parameters, the Client Layer ensures that training is both energy-efficient and aligned with NWDAF's goals for reducing overhead in real-time analytics.

2.3. Edge Layer

The Edge layer operates as an intermediary between the Client layer and higher-tier layers like the Fedge layer and Cloud layer. Its primary function is to collect and process model updates from the clients before passing them upwards in the network. The Edge layer uses NWDAF to optimize the flow of data and manage network functions such as AMF and SMF by ensuring that relevant, context-aware information is processed efficiently.

The interaction with NWDAF allows the Edge Layer to perform distributed analytics, reducing latency and network load by processing data closer to its source. For instance, by using ML models within the Edge Layer, the system can make real-time decisions about what data is important and should be sent to the Fedge layer or NWDAF for further analysis. This selective forwarding of information conserves energy and ensures that the Edge layer operates efficiently within the NWDAF framework.

Furthermore, this layer ensures that communication between client devices and the network is optimized for energy savings, avoiding redundant data aggregation. The edge nodes also facilitate secure and efficient communication of aggregated model updates from multiple clients to the Fedge ayer, aligning with NWDAF's objective of streamlining distributed analytics across the network.

2.4. Fedge Layer

The Fedge layer acts as the aggregation and coordination layer in this architecture. It serves as a crucial point for managing and optimizing the flow of data between Edge and Cloud layers. This layer aggregates model updates from multiple Edge layers, performing complex tasks such as clustering and aggregation to improve scalability and efficiency in the FL process.

In this architecture, the Fedge layer closely integrates with NWDAF to ensure that the aggregated data from various network functions (like AF, SMF, AMF, NEF, and PCF) is analyzed efficiently. As depicted in the diagram, Fedge plays a key role in facilitating horizontal (East/West) communication between NWDAF instances, enhancing network-wide coordination and distributed analytics. This coordination is crucial for managing data from different network segments and ensuring consistent and efficient learning processes across the system.

By aggregating model updates before they are sent to the Cloud layer, the Fedge layer minimizes computational redundancy and energy consumption, which aligns with the energy efficiency goals of the system. This hierarchical aggregation also prevents unnecessary re-training of models across different edges, further reducing the system's computational burden.

2.5. Cloud Layer

The Cloud layer serves as the top tier in the architecture, where global model synchronization and storage occur. It is responsible for aggregating insights from lower layers, including the Fedge Layer, and coordinating network-wide optimization strategies.

The Cloud layer interacts with NWDAF by handling centralized analytics functions and ensuring that the insights gathered from the Fedge layer are used for system-wide optimization. This interaction allows the Cloud Layer to update models and analytics functions across the network based on real-time feedback from NWDAF, ensuring that the system remains scalable and energy-efficient.

The Cloud layer focuses on maintaining model synchronization while minimizing communication overhead with the lower tiers. By optimizing when and how updates are pushed to the lower layers, the Cloud layer ensures that the system remains responsive and aligned with energy-efficient operations. Additionally, by interacting with the Network Repository Function (NRF), the Cloud Layer ensures the discoverability and integration of analytics services across the network.

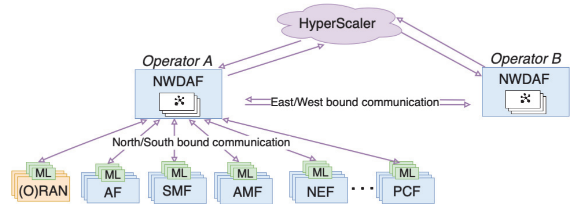

Figure 2.

Proposed multi-tier FL architecture within 3GPP

2.6. Integration with NWDAF

The integration of the multi-tier FL architecture with NWDAF is achieved through several key mechanisms, ensuring compliance with 3GPP standards and seamless operation within 5G and beyond networks.

1. **Standardized interfaces and protocols:**
   The architecture utilizes standardized interfaces defined by 3GPP, such as the NWDAF interface, to ensure compatibility and interoperability with existing network functions. This allows for efficient communication between NWDAF and other network functions, facilitating data exchange and analytics reporting.
2. **Security and privacy compliance:**
   Adhering to 3GPP security frameworks, the architecture ensures secure communication and data handling across all layers. Secure protocols are used for transmitting model updates and analytics data, protecting user privacy and data integrity. By preserving user privacy and minimizing the transmission of sensitive data, the architecture complies with data protection regulations and user expectations.
3. **Support for network slicing:**
   The hierarchical design of the architecture supports NWDAF's role in managing network slices by providing analytics tailored to specific slice requirements. Each layer can process and aggregate data relevant to particular network slices, enabling more efficient and customized network management. This capability enhances the network's ability to provide specialized services and optimizes resource allocation, aligning with 3GPP standards.
4. **Energy efficiency optimization:**
   By minimizing redundant data processing and optimizing computational resource allocation across layers, the architecture contributes to NWDAF's objectives for energy efficiency. Techniques such as adaptive training at the Client Layer, intelligent forwarding at the Edge Layer, and efficient model aggregation at the Fedge Layer collectively reduce energy consumption, which is crucial for sustainable network operations in 6G.
5. **Real-time analytics and scalability:**
   Integration with NWDAF enables real-time analytics, which is essential for applications requiring low latency and immediate responsiveness. The multi-tier design allows for scalable deployment, supporting a growing number of devices and services in 6G. This scalability aligns with NWDAF's need to accommodate increasing network complexity while maintaining high performance.

3. Results Analysis

In this section, we assess the effectiveness of the multi-tier FL architecture, emphasizing its ability to handle data heterogeneity and its implications for energy efficiency and integration with the NWDAF. The evaluation leverages results from our previous study [12], focusing on findings that are particularly pertinent to energy efficiency in 6G.

3.1. Methodology

The architecture's performance was evaluated using the Federated Averaging (FedAvg) algorithm on the MNIST dataset under non-Independent and Identically Distributed (non-IID) conditions. Two scenarios were designed to test the architecture's capability to manage data heterogeneity:

Generalisable non-IID scenario: The dataset was evenly divided among three clients, each receiving a balanced mix of all digit classes. This scenario serves as a baseline to evaluate performance under typical non-IID conditions.

Non-generalisable non-IID scenario: One client received images of only two digit classes, while another received images of eight classes, creating an extreme imbalance. This scenario challenges the architecture's ability to handle significant data skew, which is critical for real-world applications where data distributions are often uneven.

A Simple Convolutional Neural Network (SimpleCNN) model was used across all clients to maintain consistency. The primary focus was on measuring model accuracy and convergence speed, metrics that directly impact computational load and energy consumption.

3.2. Results and Analysis

In both scenarios, the multi-tier FL architecture outperformed the standard FL model in terms of convergence speed and final accuracy:

Faster Convergence: The multi-tier FL model reached higher accuracy levels in fewer training rounds compared to the standard FL model. This reduction in training rounds implies fewer communication cycles and less computational effort, leading to energy savings.

Robustness to Data Heterogeneity: In the non-generalisable non-IID scenario, the multi-tier FL architecture maintained consistent accuracy across clients, effectively mitigating the adverse effects of extreme data imbalance. This robustness is essential for NWDAF's ability to provide reliable analytics in diverse and dynamic network conditions.

These results demonstrate the architecture's efficiency in handling non-IID data distributions, a common challenge in federated networks. Detailed figures and quantitative results are provided in our previous work [12].

3.3. Implications for Energy Efficiency and NWDAF Integration

The observed improvements have significant implications for energy efficiency and NWDAF integration:

1. **Energy Efficiency:** Faster convergence and reduced training rounds translate directly into lower energy consumption. By minimizing the computational workload and the frequency of communication between clients and servers, the multi-tier FL architecture conserves energy—a critical consideration for battery-powered devices and large-scale networks in 6G.
2. **Optimized Resource Management:** The architecture employs AI-driven strategies that adapt to varying network conditions and data distributions. By focusing computational efforts where they are most needed and avoiding redundant processing, it reduces unnecessary energy expenditure.
3. **Alignment with NWDAF:** The architecture's ability to handle heterogeneous data while maintaining high accuracy aligns with NWDAF's objectives of efficient and privacy-preserving analytics. Its hierarchical structure maps onto NWDAF's distributed analytics framework, facilitating seamless integration and enhancing overall network performance.

These findings underscore the suitability of the multi-tier FL architecture for deployment in energy-sensitive and data-intensive environments characteristic of 6G.

4. Conclusion & Future Directions

In this paper, we evaluated the proposed mult-tier FL architecture for the energy efficiency demands of 6G, integrating seamlessly with the NWDAF as defined by 3GPP standards. Initial simulations demonstrated the framework's potential in reducing energy consumption while maintaining high model accuracy, even with non-IID data distributions. However, as this is a preliminary proposal, several challenges need to be addressed. Real-world evaluation is required to validate the architecture's effectiveness across diverse and complex datasets representative of 6G scenarios. Additionally, issues related to security and the heterogeneity of edge devices must be overcome to ensure robust and scalable deployment.

Future work will focus on conducting extensive trials in practical settings, refining AI algorithms for better adaptability, and testing with more complex and larger-scale datasets. Addressing these challenges is crucial for advancing towards energy-efficient, high-performance networks in the 6G era. By overcoming these obstacles, the proposed multi-tier FL architecture can significantly contribute to sustainable and intelligent 6G networks, enhancing distributed learning capabilities and optimizing network operations in compliance with industry standards.

References

[1] 3GPP, "Architecture enhancements for 5G System (5GS) to support network data analytics services," Technical Specification (TS) 23.288, 3rd Generation Partnership Project (3GPP), 09 2022, version 17.6.0.

[2] 3GPP, "Architecture enhancements for 5G System (5GS) to support network data analytics services," Technical Specification (TS) 23.288, 3rd Generation Partnership Project (3GPP), 03 2024, version 18.5.0.
[3] 3GPP, "Architecture enhancements for 5G System (5GS) to support network data analytics services," Technical Specification (TS) 23.288, 3rd Generation Partnership Project (3GPP), 09 2024, version 19.0.0.
[4] Y. Xiao, G. Shi, and M. Krunz, "Towards ubiquitous AI in 6G with federated learning," *arXiv preprint arXiv:2004.13563*, 2020.
[5] N. Hu, Z. Tian, X. Du, and M. Guizani, "An energy-efficient in-network computing paradigm for 6G," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 4, pp. 1722–1733, 2021.
[6] Z. Wang, Y. Zhao, Y. Zhou, Y. Shi, C. Jiang, and K. B. Letaief, "Over-the-Air Computation for 6G: Foundations, Technologies, and Applications," *IEEE Internet of Things Journal*, 2024.
[7] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G Internet of Things: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 359–383, 2021.
[8] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019-2019 IEEE international conference on communications (ICC)*, pp. 1–7, IEEE, 2019.
[9] L. Shanmugam, R. Tillu, and M. Tomar, "Federated learning architecture: Design, implementation, and challenges in distributed AI systems," *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, vol. 2, no. 2, pp. 371–384, 2023.
[10] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1205–1221, 2019.
[11] L. Qian, P. Yang, M. Xiao, O. A. Dobre, M. D. Renzo, J. Li, Z. Han, Q. Yi, and J. Zhao, "Distributed Learning for Wireless Communications: Methods, Applications and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 3, pp. 326–342, 2022.
[12] B. Satwat, D. Tasos, K. Kasra and I. Muddesar, "Architectural blueprint for heterogeneity-resilient federated learning," *IET 6G and Future Networks Conference (IET 6G 2024)*, London, UK, 2024, pp. 62–68, doi: 10.1049/icp.2024.2236.

Biographies

Satwat Bashir is a PhD student at London South Bank University. Her research focuses on Federated Learning within the domain of edge computing and machine learning as part of the Smart Internet Technologies (SITHub) research group. Satwat holds a Master's degree in Cognition and Computation from Birkbeck, University of London, and a BSc in Computer Engineering from the University of Engineering and Technology, Lahore.

Tasos Dagiuklas received the Engineering degree in electrical engineering from the University of Patras, Greece, in 1989, the M.Sc. degree in electrical engineering from The University of Manchester, U.K., in 1991, and the Ph.D. degree in electrical engineering from the University of Essex, U.K., in 1995. He is a leading researcher and an expert in the field of smart internet technologies. He is the leader of the Smart Internet Technologies (SITHub) Research Group, London South Bank University, where he also acts as the Head of the Cognitive Systems Research Centre (CSRC). He has been a principal investigator, a co-investigator, the project and technical manager, a coordinator, and a focal person of more than 25 international research and development and capacity training projects in the areas of fixed-mobile convergence, 5G/6G technologies, VoIP, and multimedia networking. His research interests include systems beyond 5G/6G technologies, programmable networks, UAVs, V2X communications, and cyber security for smart internet systems.

Kasra Kassai is the Associate Head of Department and a researcher in London South Bank University, specializing in sustainable edge computing, with a focus on energy optimisation. In his role, and as a Fellow of the Higher Education Academy, Kasra actively contributes to green computing and resource management research, collaborating with the Smart Internet Technologies (SITHub) research group and the Cognitive Systems Research Centre (CSRC). They mentor students and emerging scholars, promoting innovative solutions in sustainable edge computing. Kasra holds a BSc, and an MSc in Software Engineering from City, University of London, with expertise in managing latency in heterogeneous networks and experience coordinating EU-funded projects. He is also a member of IEEE, BCS, and ACM.

Muddesar Iqbal received the Ph.D. degree from Kingston University, U.K., in 2010. He is currently a Senior Lecturer in mobile computing with the Division of Computer Science, School of Engineering. He co-invented four patented inventions in the area of the IoT, intelligent systems, and 5G enabling technologies. He published nearly 100 peer-reviewed articles in reputed journals, conference proceedings, and book collections. His research interests include 6G, Internet of sense for industry 5.0, and collaborative cognitive communication systems social and intelligent autonomous machines/robotics. His Ph.D. was funded by EPSRC Doctoral Training Award for 4G-based Reconfigurable Mobile Healthcare System. He has won more than 15 research and development and capacity building funding Grants from different national and international funding agencies.