

Paper Type: Original Article

Digital Watermarking and Federated Learning: A Synergistic Approach

Asmaa Alqassab ^{1,*} 

¹ Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul, Nineveh, Iraq;
asmaa_mow@uomosul.edu.iq.

Received: 05 Dec 2024

Revised: 15 Feb 2025

Accepted: 31 May 2025

Published: 03 Jun 2025

Abstract

This work investigates the collaborative relations between Digital Watermarking and Federated Learning. Digital watermarking has the major part in data integrity and content rights enforcement. As another option, as part of the effort to ensure privacy, federated learning put forth by Google in 2016, deals with the difficulties of the data being distributed, which makes it an effective strategy in diverse fields. Accordingly, this paper explores how to put those techniques together such that they complete each other. It considers the federated learning's possible utilization in an effort to refine the performance of the watermarking algorithm. Moreover, this work emphasizes the watermarking role in improving federated learning for securing intellectual rights, by means of reviewing existing strategies.

Keywords: Watermarking; Federated Learning; Distributed Data; Cybersecurity; Data Protection.

1 | Introduction

Federated Learning offers faster response time, energy conservation, and smarter models, at the same time providing privacy. It is a machine learning setting in which the training data keeps distributed among a great many of clients after a top-quality centralized paradigm being trained, each with somewhat slow and undependable network connectivity [1]. Suggested approach proffers one more direct advantage, that's the enhanced design on your own mobile may be employed promptly, powerful experiences tailored by the manner you utilize your mobile along with supplying the shared model with updates [2].

Federated learning holds significant promise across an array of industries, including not only sales and finance but also computer vision [3], IoT [4], environment protection [5] and in electric vehicle networks [6]. In these sectors, the amalgamation of data for machine learning model training is often impeded by critical factors such as protection intellectual property, adhering to stringent privacy regulations, and fortifying data security measures [7, 8].

In this paper, we study the relationship between federated learning and watermarking, which is a technique utilizes to embed imperceptible and unique information, or "watermarks," into digital medium such as video, audio, or images. These watermarks serve multiple intentions, involving copyright protection, tracking digital



Corresponding Author: asmaa_mow@uomosul.edu.iq



<https://doi.org/10.61356/j.nois.2025.6610>



Licensee **Neutrosophic Optimization and Intelligent systems**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

contents, meta data saving, and ownership verification. They are designed to be resistant to removal or alteration, ensuring the integrity and traceability of the digital content [9, 10].

This work will explore the intersection between federated learning and watermarking, spotlighting two key areas:

- a) Preserving Participant Copyrights: aims to provide insight on the relation between those two zones, in which watermarking plays as a defender of ownership in the cooperative world of FL.
- b) Enhancing Watermarking Algorithms: seeks to fortifying the bond between those two zones, moreover, how FL mechanisms may qualify the production of more influential and sturdy watermarking approaches.

2 | Theoretical Background

In this section, a theoretical background explaining the federated learning and watermarking is presented.

2.1 | Federated Learning

In recent years, an exponential evolution in industrial engineering of deep learning, machine learning, smart production and artificial intelligence. However; there's a big concern about who gets to use your personal data. There are rules, like GDPR "General Data Protection Regulation", that say only you have control over your data, and companies can't use it without your permission. Secondly, these industries have a hard time-sharing important data that helps make things work better. It's like they keep their data in separate rooms, and this makes it tough to improve how machines learn and work together.

So, federated learning presented by Google in 2016 [11], is tend to solve the issue of data disclosure, by allowing each participant to share their model update only without the data.

To explain the federated learning, imagine you have a big software development project, and you want to build a highly specialized program. Instead of having all the programmers working in one central office, you decide to hire several programming teams located in different parts of the world. Each programming team has its own data and code that they're working on. They keep their data private because it contains valuable information and algorithms related to their specific task. Similarly, in federated learning, different entities (like users or organizations) have their own data, and they want to keep it private due to reasons like privacy protection, intellectual property rights, or data security.

Federated learning allows different entities to collaborate on training a shared machine learning model without centrally aggregating their data [12]. In both cases, the goal is to achieve a common objective without revealing the individual components (data or code) to each other. The final software product in our analogy represents the machine learning model in federated learning. Each programming team (or entity in federated learning) contributes their expertise to build a better end product (or model) collectively.

Federated learning is a machine learning approach where the model is trained across decentralized devices or servers holding local data samples, and the training process occurs without exchanging raw data [13].

The following steps explain the process of federated learning [8]:

1. Every mobile loads a public paradigm for the next local training.
2. Loaded common paradigm enhanced with numerous local updates together with local data, from diverse smartphones individually, afterwards and in an encryption pattern, associated gradient information is uploaded to the cloud.
3. The update of local paradigms carried out within cloud, is forwarded to the mobile as a regenerated global paradigm.

4. Reiterate the preceding steps till the time limit meets or the paradigm accomplishes a specific required performance.

Figure 1 explains the work of federated learning.

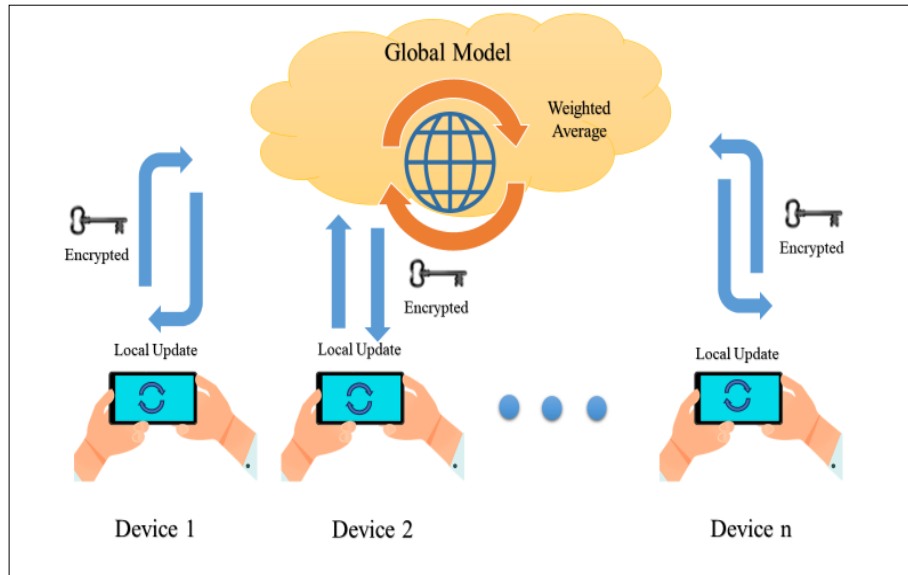


Figure 1. Federated Learning Updates.

A key challenge in the federated learning lies in determining the rightful contributors and owners of specific models. This challenge arises from the fact that, in a collaborative learning environment, it may be unclear who has made a certain contribution to a particular model. Consequently, it is important to protect the intellectual property and ensure proper recognition for the work of contributors [14].

2.2 | Digital Watermarking

Digital watermarking, a process involving the insertion of a watermark signal, whether in the form of text or a small binary image, into a target host digital content, serves as the foundation for various applications as medical image exchange [15], Electrocardiogram signal security [16], and copyright protection [17].

A digital watermarking system comprises two main phases: embedding, where the watermark is integrated into the host digital content, and extraction, where the watermark is retrieved from the host digital content. Extraction can be categorized into two types: blind extraction, where the original content is not needed for retrieval, and non-blind extraction, where the host digital content is essential for extraction.

The key elements of a watermarking process encompass the following components: the host (original) digital content, which is the target for watermarking; the watermark, representing the information to be embedded, which could be a company logo or metadata; the key, an optional encryption key used to encrypt the watermark for enhanced security; and the watermarked digital content, the resulting content that implicitly contains the watermark and conveys the intended information or identification [9, 18].

Figures 2 and 3 shows the embedding and extracting processes, respectively. If the technique used for digital watermarking depends on the original data being present in order to extract the embedded watermark, it is considered non-blind extraction. On the other hand, with blind watermarking methods, the extraction procedure does not require the original data. This distinction is shown in the representation by the existence or absence of a connection to the original data, which is commonly shown as dotted lines in Figure 3.

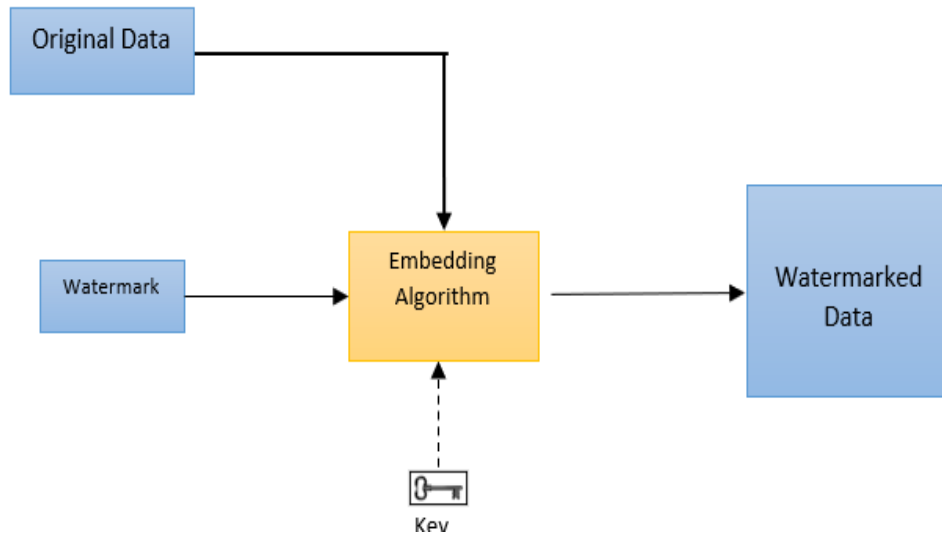


Figure 2. Watermark Embedding.

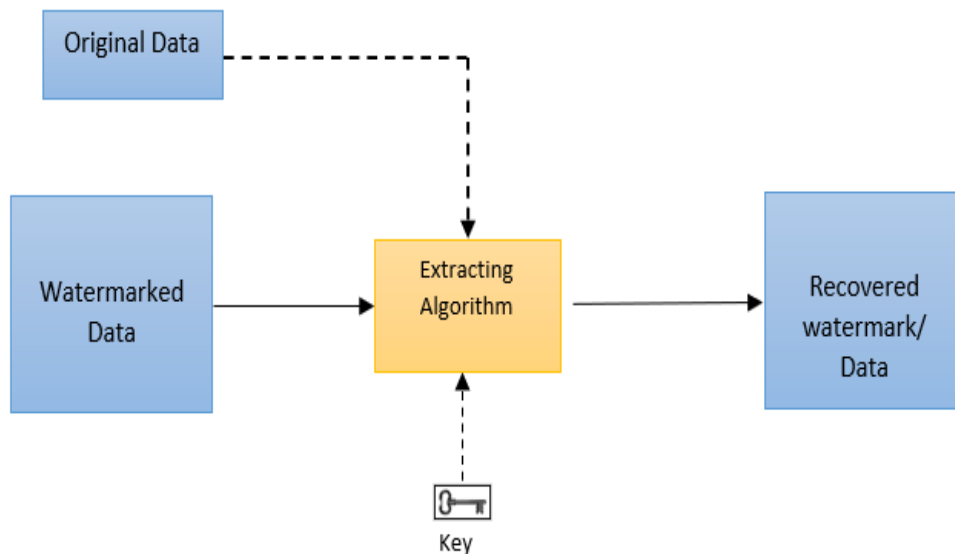


Figure 3: Watermark Extracting.

Digital watermarking serves as a powerful tool in copyright protection. It enables content creators and owners to assert their ownership and protect their intellectual property rights. These digital watermarks can contain information about the copyright holder, usage rights, or other metadata. When someone tries to use, share, or distribute the content, watermarks facilitate legal recourse if copyright violations occur. In this way, digital watermarking plays a crucial role in solving copyright for contributors in federated learning. In the same time, federated learning can be used to train a network to find the best embedding method for a watermark. The following section will explain the possibilities of cooperation between federated learning and watermarking.

2.3 | Cooperative Relationship between Federated Learning and Watermarking

In this section, the opportunities that arise from the combination of Federated Learning (FL) and watermarking techniques will be explored. the objective is to investigate the latest advancements in this field, presenting the advantages that FL offers to watermarking and also, the benefits that watermarking brings to FL.

2.4 | The Role of watermarking in FL copyright protection

Watermarking is primarily concerned with embedding and detecting digital information within multimedia content. Integrating watermarking into FL can enhance the traceability and authenticity of data shared among

participating devices or users. This added layer of security can be particularly critical for many applications, as the applications where source authentication is important, e.g. intellectual property protection.

An attempt is in [19], in this paper, the authors have introduced a novel client-side FL watermarking scheme tailored to address copyright protection concerns within the secure FL framework that employs homomorphic encryption (HE). The primary motivation for their work arises from the potential risks of unreliable participants gaining access to jointly trained models in FL. Also, Application of homomorphic encryption (HE) in secure FL framework prevents the central server from accessing plaintext models. Thus, it is no longer feasible to embed the watermark at the central server using existing watermarking schemes. To mitigate this, they propose a unique approach that embeds watermarks directly into models on the client-side, which was previously challenging due to the use of HE.

Paper [20] rolls out Merkle-Sign, a watermarking framework that blends cryptographic techniques with advanced watermarking design for distributed storage setups. The focus here isn't just security in the abstract it's about proving ownership in federated learning (FL) systems where data and models are spread across different parties. To make the system less theoretical and more workable, the authors add another layer called the Autoencoder-based Trigger Generator for Federated Learning (ATGF). It acts like a practical tool that keeps the watermark adaptable to messy, real-world conditions. The experiments don't just tick boxes they actually show that the framework holds its ground, keeping the watermark intact and the model's authorship verifiable even in distributed environments.

In Paper [21], the researchers come up with WAFFLE, a creative attempt to settle the long-standing question of who owns what in a client-server federated learning setup. The idea is tested on two classic datasets CIFAR-10 and MNIST and the results point to a system that's not only accurate but steady across different settings. To push things further, the paper adds WAFFLEPATTERN, a data-independent watermarking approach made specifically for deep neural network image classifiers. Instead of relying on the quirks of training data, it builds watermarks that stay stable and effective regardless of the dataset, which is a big deal for generalization.

Paper [22] takes a step back to look at the bigger picture. It unpacks how federated learning and AI are slowly reshaping the conversation around data privacy, legal responsibility, and the limits of centralized machine-learning systems. The argument is that FL isn't just another algorithmic tweak it's a kind of structural rethink of how intelligence can be built collaboratively without giving up control of data.

Finally, Paper [23] explores a more defensive angle: how to embed watermarks into FL models to assert ownership and protect intellectual property. The proposal acts like an audit mechanism a way to verify authorship even after a model has been shared or fine-tuned elsewhere. The authors see this as part of a larger shift toward secure and lawful AI collaboration, where trust doesn't rely on blind faith but on verifiable proof built into the system itself.

The key innovation in this method is the design of a backdoor-based watermarking scheme. This scheme allows model owners to embed predefined noise patterns into FL deep models, effectively providing robust copyright protection. Importantly, this watermarking process occurs on the client-side, ensuring that sensitive gradient information is kept private and inaccessible to the central node. The proposed work maintains FL model performance and show the robustness of the watermarking technique in the face of various challenges.

Table 1 abbreviates the mentioned attempts, however a review for different attempts was mentioned in [24].

Table1. List of Watermarking in FL.

NO.	Year	Ref No.	Main Focus	Notable Features
1	2022	[19]	Copyright protection in secure FL	Embedding watermarks on the client-side; Addressing risks in FL; Use of homomorphic encryption (HE)
2	2021	[20]	Ownership verification using watermarks	Incorporates advanced watermarking schemes ; Introduces ATGF watermarking scheme
3	2021	[21]	Ownership protection in client-server FL	Embedding backdoor-based watermarks; Tested on MNIST and CIFAR10 datasets
4	2023	[22]	Data privacy and legal challenges in FL	Provides an overview of AI and FL; Introduces FedIPR for ownership verification
5	2022	[23]	Copyright protection in FL	Embeds predefined noise patterns in FL models on the client-side; Maintains FL model performance

2.5 | The Role of Federated Learning in Watermarking Systems

The distributed nature of Federated Learning allows multiple entities to collaborate in improving a shared model while preserving data privacy and security. In the context of watermarking, FL offers a promising avenue for enhancing the robustness and effectiveness of watermarking algorithms. By using the power of FL, watermarking schemes can potentially benefit from using several models to be trained while protecting trained data.

An example of that, a paper [25], introduces a novel approach, a robust zero-watermarking scheme based on federated learning. In this scheme, a sparse autoencoder network is trained using federated learning. The trained network is then employed to extract essential image features from dermatology medical images. These extracted features undergo a two-dimensional Discrete Cosine Transform (2D-DCT) process to select low-frequency transform coefficients for the creation of zero-watermarks. The experimental results presented in the paper shows superior performance of the proposed scheme schemes, in scenarios involving conventional attacks and geometric attacks. Importantly, this alters the crucial information contained within the medical images nor compromises the confidentiality of patient data.

The utilization of federated learning in watermarking is indeed an emerging area with promising potential, but currently, there are relatively few research efforts in this direction. This observation highlights an important research gap and the need for increased exploration and innovation in this domain. In the next section, the importance of federated learning in watermarking is presented.

3 | Methodology

In this section, we explore opportunities for enhancing the field of Federated Learning (FL) through the integration of watermarking techniques, and we also examine how watermarking can benefit from the principles of federated learning.

3.1 | Enhancing Watermarked Federated Learning Approaches

The following suggestions can be applied to enhance federated learning systems by using watermarking approaches:

- **Adaptive Watermarking:** Develop watermarking schemes that adapt dynamically based on the level of trustworthiness of participants. Participants with a higher trust level may be assigned more critical watermarking, while others receive a lighter watermark.
- **Scalable Watermarking:** Address the scalability challenges in federated learning watermarking, especially as the number of participants increases. Research can study how to maintain efficiency and security in large-scale federated learning settings.

- Legal Considerations: Conduct research on the legal and ethical implications of watermarking in federated learning. Examine the admissibility of watermarked models as evidence in legal proceedings and explore the ethical guidelines surrounding watermarking.
- Real Time watermarking: Applying privacy real time federated learning, for instance Field Programmable Gate Arrays (FPGA) has been used in many applications of federated learning [26, 27], watermark can be applied in these devices to secure the system in real time.

3.2 | Enhancing Watermarking Algorithms using FL

In situations where multiple content creators or owners want to watermark their media but do not want to share the original content, federated learning can be employed. Each owner can embed their watermark locally on their media using their private keys or watermarking algorithms without sharing the original content. Also, Federated learning can enable watermarking techniques that adapt to changing conditions. Content owners can collectively train models that adjust watermarking parameters based on the type of content, distribution channels, or security requirements.

For instance, in image processing FL can be used to be performed locally on individual client devices. This means that the raw image data remains on the client side and is never shared centrally. Each participant can embed a unique watermark into their local images before contributing to the global model. This watermark serves as a digital signature, confirming the data's origin and ownership. Image and video. The algorithms can be enhanced with federated learning, as each client can use different images with different characteristics, these will enhance the embedding algorithms to achieve the best invisibility and robustness against attacks; While in the same time keep the privacy of user's data. There is currently limited research in the application of federated learning to image, video, and text watermarking. However, these areas hold promise as potential avenues for future research in the field. Figure 4 illustrates the suggested watermarked model.

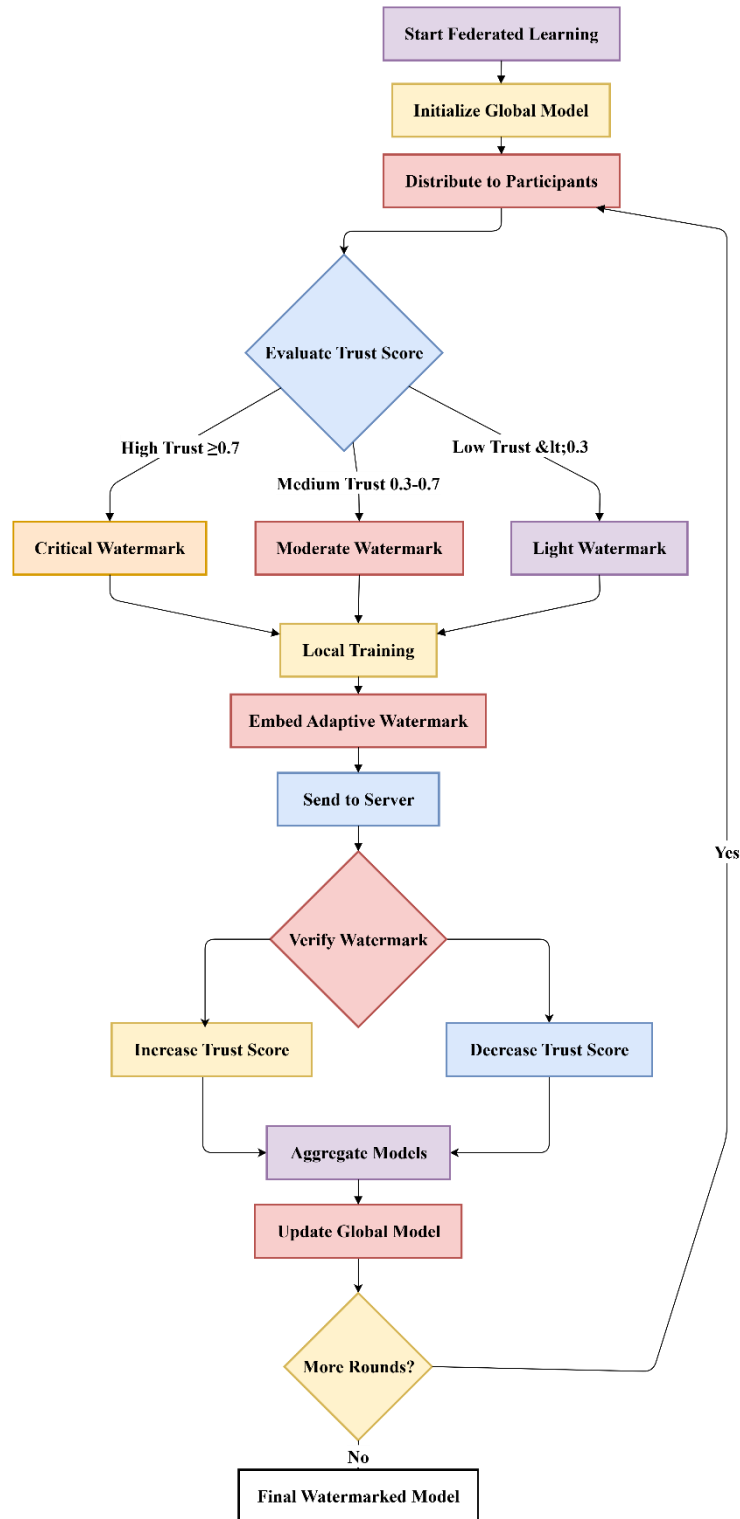


Figure 4. The suggested watermarked model

4 | Conclusion

In conclusion, this paper has illuminated the symbiotic relationship between Federated Learning (FL) and Digital Watermarking, highlighting the significant benefits that can be derived from their integration. A literature review shows the role of watermarking in enhance the traceability and authenticity of data shared among participating devices or users in federated learning. However, some areas of development can be used in future studies as adaptive watermarking, scalability and explore real-time applications. In the other hands,

the number of researches in utilizing federated learning is very limited, and there is a big gap in research area to use federated learning to produce more efficient watermarking algorithm. For instance, in image watermarking, federated learning can be used to train the model with different users without revealing the personal images.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," arXiv preprint arXiv:1610.05492, 2016.
- [2] B. McMahan and D. Ramage, "Federated learning: Collaborative machine learning without centralized training data," Google Research Blog, p. 3, 2017.
- [3] X. Han, H. Yu, and H. Gu, "Visual inspection with federated learning," in Image Analysis and Recognition, F. Karray, A. Campilho, and A. Yu, Eds. Cham: Springer, 2019, pp. 52–64. doi: 10.1007/978-3-030-27272-2_5.
- [4] U. M. Aïvodji, S. Gambs, and A. Martin, "IOTFLA: A secured and privacy-preserving smart home architecture implementing federated learning," in Proc. IEEE Security and Privacy Workshops (SPW), 2019, pp. 175–180. doi: 10.1109/SPW.2019.00041.
- [5] B. Hu, Y. Gao, L. Liu, and H. Ma, "Federated region-learning: An edge computing based framework for urban environment sensing," in Proc. IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1–7. doi: 10.1109/GLOCOM.2018.8647649.
- [6] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, M. D. Mueck, and S. Srikanteswara, "Energy demand prediction with federated learning for electric vehicle networks," in Proc. IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1–6. doi: 10.1109/GLOBECOM38437.2019.9013587.
- [7] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Trans. Intell. Syst. Technol. (TIST), vol. 10, no. 2, pp. 1–19, 2019.
- [8] L. Li, Y. Fan, M. Tse, and K. Y. Lin, "A review of applications in federated learning," Computers & Industrial Engineering, vol. 149, 106854, 2020.
- [9] I. Cox, M. Miller, J. Bloom, and C. Honsinger, "Digital watermarking," J. Electron. Imaging, vol. 11, no. 3, pp. 414–414, 2002.
- [10] M. Begum and M. S. Uddin, "Digital image watermarking techniques: A review," Information, vol. 11, no. 2, 110, 2020.
- [11] Google, "Federated Learning," [Online]. Available: <https://federated.withgoogle.com/>. [Accessed: Sep. 23, 2023].
- [12] M. Moshawrab, M. Adda, A. Bouzouane, H. Ibrahim, and A. Raad, "Reviewing federated learning aggregation algorithms: strategies, contributions, limitations and future perspectives," Electronics, vol. 12, no. 10, 2287, 2023.
- [13] "Federated Learning," RapidMiner Glossary, [Online]. Available: <https://rapidminer.com/glossary/federated-learning/>. [Accessed: Sep. 23, 2023].
- [14] Q. Yang et al., "Federated learning with privacy-preserving and model IP-right-protection," Mach. Intell. Res., vol. 20, no. 1, pp. 19–37, 2023.
- [15] A. H. Allaf and M. A. Kbir, "A review of digital watermarking applications for medical image exchange security," in Innovations in Smart Cities Applications Edition 2: Proc. Third Int. Conf. Smart City Applications, Cham: Springer, pp. 472–480, 2019.

- [16] A. Khaldi, M. R. Kafi, and B. Meghni, "Electrocardiogram signal security by digital watermarking," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 10, pp. 13901–13913, 2023.
- [17] T. B. Taha, R. Ngadiran, and P. Ehkan, "Adaptive image watermarking algorithm based on an efficient perceptual mapping model," *IEEE Access*, vol. 6, pp. 66254–66267, 2018.
- [18] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999, doi: 10.1109/5.771066.
- [19] W. Yang et al., "Watermarking in secure federated learning: A verification framework based on client-side backdooring," *arXiv preprint arXiv:2211.07138*, 2022.
- [20] F. Q. Li, S. L. Wang, and A. W. C. Liew, "Towards practical watermark for deep neural networks in federated learning," *arXiv preprint arXiv:2105.03167*, 2021.
- [21] B. G. Tekgul, Y. Xia, S. Marchal, and N. Asokan, "Waffle: Watermarking in federated learning," in *Proc. 40th Int. Symp. Reliable Distributed Systems (SRDS)*, pp. 310–320, 2021.
- [22] Q. Yang et al., "Federated learning with privacy-preserving and model IP-right-protection," *Mach. Intell. Res.*, vol. 20, no. 1, pp. 19–37, 2023.
- [23] B. Li, L. Fan, H. Gu, J. Li, and Q. Yang, "FedIPR: Ownership verification for federated deep neural network models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4521–4536, 2022.
- [24] M. Lansari, R. Bellafqira, K. Kapusta, V. Thouvenot, O. Bettan, and G. Coatrieux, "When federated learning meets watermarking: A comprehensive overview of techniques for intellectual property protection," *arXiv preprint arXiv:2308.03573*, 2023.
- [25] B. Han, R. Jhaveri, H. Wang, D. Qiao, and J. Du, "Application of robust zero-watermarking scheme based on federated learning for securing the healthcare data," *IEEE J. Biomed. Health Inform.*, 2021.
- [26] Z. Wang et al., "PipeFL: Hardware/software co-design of an FPGA accelerator for federated learning," *IEEE Access*, vol. 10, pp. 98649–98661, 2022.
- [27] Z. Yang, S. Hu, and K. Chen, "FPGA-based hardware accelerator of homomorphic encryption for efficient federated learning," *arXiv preprint arXiv:2007.10560*, 2020.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.