International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **02** | **2026**

A Comprehensive Review on Federated Learning: Architectures, Aggregation, Security, Heterogeneity and Future Directions

Mr. Yogesh Bhikusing Jadhao[1]
[1] *Research Scholar, Computer Science and Engineering, SKU University, MP, India*

DOI: 10.5281/zenodo.19284806

ABSTRACT

Federated Learning (FL) has emerged as a promising distributed machine learning paradigm that enables multiple clients to collaboratively train a shared model without exchanging raw data. This approach addresses growing concerns related to data privacy, security regulations, and communication efficiency in modern intelligent systems. Unlike traditional centralized learning, FL performs local model training at edge devices and only transmits model updates to a central coordinator for aggregation. However, practical deployment of FL faces several challenges, including statistical heterogeneity, system heterogeneity, privacy leakage, security attacks, fairness issues, and inefficient aggregation. This paper presents a comprehensive and up-to-date review of federated learning by systematically analyzing its architecture, aggregation strategies, heterogeneity handling techniques, security and privacy mechanisms, and fairness-aware learning methods. Recent advances in secure aggregation, differential privacy, homomorphic encryption, and robust optimization are discussed. Furthermore, open research challenges and future directions such as decentralized FL, scalable training, and trustworthy FL systems are highlighted. This survey aims to provide researchers and practitioners with a consolidated understanding of the state-of-the-art developments and emerging trends in federated learning.

Keywords:- Federated Learning, Privacy-Preserving Machine Learning, Secure Aggregation, Heterogeneous Data, Distributed Optimization, Edge Intelligence.

1. INTRODUCTION

The rapid growth of data-driven applications in healthcare, smart cities, finance, Internet of Things (IoT), and mobile services has significantly increased the demand for large-scale machine learning models. Conventional centralized learning frameworks require collecting massive volumes of data from distributed sources into a single server for training. However, such a paradigm raises serious concerns regarding user privacy, data ownership, communication overhead, and compliance with data protection regulations such as GDPR and HIPAA.

Federated Learning (FL) has been proposed as an effective solution to overcome these limitations by enabling collaborative model training while keeping sensitive data locally at participating clients. In an FL system, each client trains a local model using its private dataset and shares only the model parameters or gradients with a central server, which aggregates them to produce a global model. This decentralized training paradigm reduces the risk of data leakage and supports large-scale learning across geographically distributed and resource-constrained devices.

Despite its advantages, FL introduces several fundamental challenges. The data across clients are typically non-independent and non-identically distributed (non-IID), leading to slow convergence and degraded accuracy. Clients may also have diverse computational capabilities, network conditions, and model architectures, resulting in system and model heterogeneity. Moreover, recent studies have shown that model updates can still leak sensitive information through gradient inversion, membership inference, and poisoning attacks. Fairness and incentive issues further complicate the collaborative learning process.

This paper provides a comprehensive survey of recent research efforts in federated learning, focusing on architectural design, aggregation optimization, heterogeneity mitigation, security and privacy preservation, and fairness-aware learning. The contributions of this review are summarized as follows:

A systematic overview of the federated learning architecture and training workflow.
A detailed classification of aggregation techniques and optimization strategies.
An in-depth analysis of data, model, and system heterogeneity in FL.
A review of major security threats and corresponding defense mechanisms.
A discussion of fairness, applications, and future research directions.

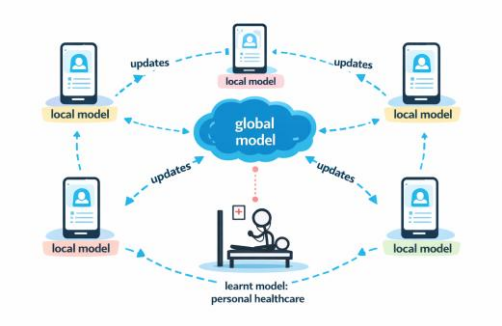

Fig. 1. An example of the FL pipeline.

2. FEDERATED LEARNING ARCHITECTURE

A typical federated learning system consists of a central server and a large number of distributed clients, such as mobile devices, edge nodes, or organizations. The overall training process is iterative and proceeds through multiple communication rounds. In each round, the server broadcasts the current global model to a subset of selected clients. Each client performs local training using its private dataset and sends the updated model parameters back to the server. The server then aggregates these updates to generate a new global model. Mathematically, let there be N clients, each holding a local dataset Di. The objective of FL is to minimize a global loss function:

$$\min_{w} F(w) = \sum_{i=1}^{N} p_i F_i(w),$$

where w denotes the model parameters, $Fi(w)$ is the local loss function at client i, and pi represents the relative importance of each client, often proportional to its data size.
The standard FL pipeline includes the following steps:
Client Selection: A subset of clients is selected in each communication round based on availability and resource constraints.
Local Training: Each selected client performs several epochs of training on its local data.
Model Upload: Clients send encrypted or compressed updates to the server.
Aggregation: The server aggregates the received updates using a predefined strategy.
Model Distribution: The updated global model is redistributed to clients for the next round.
This process continues until convergence or a predefined stopping criterion is met.

3. RELATED WORK

Early studies on federated learning focused on basic optimization and communication efficiency. The seminal work on FedAvg demonstrated that simple weighted averaging of local model parameters could achieve competitive performance compared to centralized training. Subsequently, extensive research has explored improvements in convergence speed, robustness, personalization, and privacy preservation.
Several survey papers have reviewed FL from different perspectives, including system design, communication efficiency, security, and applications. However, many existing surveys either focus on a specific sub-area or do not cover the latest advances in secure and fair FL. This paper extends prior surveys by providing an integrated view of aggregation optimization, heterogeneity handling, security defenses, and fairness mechanisms under a unified framework.

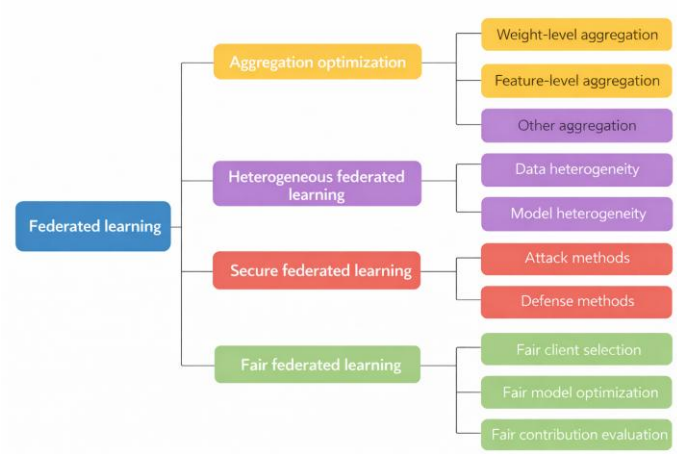

Fig. 2. Different federated learning methods.

4. AGGREGATION TECHNIQUES IN FEDERATED LEARNING

Aggregation is a core component of federated learning, as it determines how local updates are combined to form a global model. The most widely used method is Federated Averaging (FedAvg), where the server computes a weighted average of local model parameters. Although simple and efficient, FedAvg may suffer from slow convergence and instability under non-IID data distributions.

To address these issues, several advanced aggregation strategies have been proposed:

4.1 Weighted and Adaptive Aggregation

Adaptive methods assign dynamic weights to clients based on factors such as data quality, gradient divergence, and training reliability. These methods aim to reduce the impact of noisy or adversarial updates.

4.2. Robust Aggregation

Robust aggregation techniques, such as median-based and trimmed-mean methods, are designed to tolerate Byzantine or malicious clients. They improve resilience against poisoning and outlier updates.

4.3. Personalized Aggregation

Personalized FL approaches maintain both a global model and client-specific models, enabling better adaptation to local data distributions.

4.4. Knowledge Distillation-Based Aggregation

Instead of averaging parameters, some methods aggregate knowledge representations or logits, allowing heterogeneous model architectures to participate in training.

5. CHALLENGES OF HETEROGENEITY

Federated learning must operate under three major types of heterogeneity:

1. **Statistical Heterogeneity:** Non-IID and unbalanced data across clients.
2. **Model Heterogeneity:** Different model architectures and sizes.
3. **System Heterogeneity:** Varying computation, storage, and communication capabilities.

These factors significantly affect convergence behavior and system efficiency. Advanced techniques such as clustering, meta-learning, multi-task learning, and transfer learning have been proposed to alleviate these challenges.

6. SECURE AND PRIVACY-PRESERVING FEDERATED LEARNING

Although FL avoids direct data sharing, model updates may still reveal sensitive information. Attacks such as gradient inversion, membership inference, backdoor insertion, and model poisoning pose serious threats. To counter these risks, various defense mechanisms have been developed, including:

- Differential Privacy
- Secure Multi-Party Computation
- Homomorphic Encryption
- Trusted Execution Environments

7. SECURITY AND PRIVACY IN FEDERATED LEARNING

Although federated learning avoids direct sharing of raw data, recent studies have demonstrated that sensitive information can still be inferred from exchanged model updates. Hence, FL systems are vulnerable to various security and privacy attacks.

7.1. Privacy Attacks

1. **Gradient Inversion Attacks:** Attackers can reconstruct training samples by exploiting gradients shared during model updates. Optimization-based and generative reconstruction methods have shown that high-fidelity images and text can be recovered from gradients, posing a serious privacy risk.
2. **Membership Inference Attacks:** These attacks determine whether a specific data sample participated in training. Such attacks exploit overfitting characteristics of models and are particularly harmful in medical and financial datasets.
3. **Model Poisoning and Backdoor Attacks:** Malicious clients can manipulate model updates to introduce hidden backdoors that cause incorrect predictions when specific triggers appear.
4. **Model extraction Attacks:** Model extraction attacks occur when an attacker seeks to replicate or imitate a machine learning model by leveraging exposed information, including prediction responses or exchanged model updates. Such attacks compromise the secrecy of the model, risk the loss of proprietary knowledge, and weaken the security of the system, particularly in collaborative frameworks such as federated learning.

7.2. Defense Mechanisms

1. **Differential Privacy (DP):** Noise is added to model updates before sharing, ensuring that individual data contributions cannot be inferred. DP provides strong theoretical privacy guarantees but may reduce model accuracy.
2. **Secure Multi-Party Computation (SMC):** SMC enables encrypted aggregation without revealing individual updates. Secure aggregation protocols prevent the server from observing local models.
3. **Homomorphic Encryption (HE):** HE allows computation over encrypted parameters, enabling aggregation without decryption. Though secure, HE introduces heavy computational overhead.
4. **Trusted Execution Environments (TEE):** TEE-based solutions perform sensitive computations inside secure hardware enclaves, protecting model updates from adversaries even if the server is compromised.

8. FAIRNESS IN FEDERATED LEARNING

Fairness ensures that the global model does not disproportionately favor certain clients or data groups.

A. Client Fairness

Unequal participation and resource heterogeneity may cause biased learning. Techniques such as adaptive client selection, contribution-aware aggregation, and reputation-based scheduling aim to balance client influence.

B. Model Fairness

Fair FL optimizes not only global accuracy but also minimizes performance disparity across clients. Multi-objective optimization and personalized FL approaches ensure equitable accuracy among participants.

C. Incentive Mechanisms

Game-theoretic and Shapley-value based reward schemes encourage honest participation and prevent free-riding, improving fairness and sustainability.

9. APPLICATIONS OF FEDERATED LEARNING

Federated learning has been successfully applied in various domains:

1. **Healthcare:** Collaborative disease diagnosis, medical image analysis, and electronic health record modeling while preserving patient privacy.
2. **Internet of Things (IoT):** Edge-based anomaly detection, smart grid monitoring, and intelligent transportation systems.

3. **Finance:**
Fraud detection, credit scoring, and risk assessment across multiple institutions without sharing sensitive customer data.
4. **Smart Cities:**
Traffic prediction, pollution monitoring, and energy optimization using distributed urban data.
5. **Natural Language Processing:**
Keyboard prediction, speech recognition, and personalized recommendation systems.

10. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

1. **Scalability:**
Efficient training with millions of heterogeneous devices remains a challenge.
2. **Communication Efficiency:**
Gradient compression, quantization, and asynchronous updates are required for low-bandwidth environments.
3. **Decentralized Federated Learning:**
Eliminating the central server can improve robustness and trustworthiness.
4. **Explainability and Trust:**
Interpretable FL models are essential for sensitive applications such as healthcare and law.
5. **Cross-Domain and Cross-Task FL:**
Future systems should support heterogeneous tasks and dynamic environments.

11. CONCLUSION

This paper presented a comprehensive review of federated learning, covering its architecture, aggregation strategies, heterogeneity challenges, security and privacy issues, fairness considerations, and practical applications. While FL offers a promising solution for privacy-preserving collaborative learning, several open challenges remain, including scalability, robustness, and trustworthiness. Future research should focus on developing adaptive, secure, and decentralized FL systems that can operate efficiently in real-world large-scale environments.

12. REFERENCES

[1]. H. B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proc. AISTATS*, 2017.
[2]. Q. Yang et al., "Federated Machine Learning: Concept and Applications," *ACM TIST*, 2019.
[3]. K. Bonawitz et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," *ACM CCS*, 2017.
[4]. L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," *NeurIPS*, 2019.
[5]. R. Shokri et al., "Membership Inference Attacks Against Machine Learning Models," *IEEE S&P*, 2017.
[6]. N. Kairouz et al., "Advances and Open Problems in Federated Learning," *Foundations and Trends in ML*, 2021.
[7]. T. Li et al., "Fair Resource Allocation in Federated Learning," *ICLR*, 2020.
[8]. Y. Zhang et al., "Federated Learning with Differential Privacy," *ACM CCS*, 2020.
[9]. J. Sun et al., "Secure Federated Learning via Homomorphic Encryption," *IEEE TIFS*, 2021.
[10]. P. Kairouz et al., "On the Privacy of Federated Learning," *JMLR*, 2021.
[11]. M. Mohri et al., "Agnostic Federated Learning," *ICML*, 2019.
[12]. T. D. Nguyen et al., "Poisoning Attacks on Federated Learning," *IEEE TDSC*, 2020.
[13]. S. Caldas et al., "LEAF: A Benchmark for Federated Settings," *arXiv*, 2018.
[14]. Y. Li et al., "Federated Optimization in Heterogeneous Networks," *IEEE JSAC*, 2020.
[15]. A. Ghosh et al., "Robust Aggregation in Federated Learning," *AAAI*, 2022.
[16]. J. Konečný et al., "Federated Learning: Strategies for Improving Communication Efficiency," *arXiv*, 2016.
[17]. X. Wu et al., "Privacy and Security in Federated Learning: A Survey," *IEEE Access*, 2021.
[18]. S. Ramaswamy et al., "Personalized Federated Learning," *NeurIPS*, 2020.
[19]. Y. Liu et al., "Clustering-Based Federated Learning," *IEEE TNNLS*, 2022.
[20]. A. Geyer et al., "Differentially Private Federated Learning: A Client Level Perspective," *NeurIPS*, 2017.