

A Comprehensive Investigation of Federated Learning Frameworks: Architectures, Security Challenges and Future Directions

Chuanfang Wang

Portland Institute/Electronic Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, China

p23000807@njupt.edu.cn

Abstract. With the rapid development of Artificial Intelligence (AI) and Machine Learning (ML) technologies, federated learning models have been widely applied in various fields such as financial risk management, healthcare, and intelligent recognition. This study primarily analyzes the architectural characteristics, workflow, advantages, and disadvantages of three classical federated learning frameworks, namely horizontal federated learning, vertical federated learning, and federated transfer learning to identify the most suitable application scenarios, potential risks, and corresponding security solutions for each model. Furthermore, the paper discusses several major challenges commonly faced by different federated learning models, including issues such as regulatory compliance and cross-domain interoperability. Based on the analytical findings, it also explores potential future development directions of federated learning, such as the maturation of a trustworthy federated AI ecosystem and governance. Finally, the study concludes that implementing multi-layer privacy protection strategies, encrypted gradient exchange mechanisms and other methods can effectively reduce security risks in federated learning systems.

Keywords: Federated learning, data privacy and security, Horizontal Federated Learning (HFL), Vertical Federated Learning (VFL), Federated Transfer Learning (FTL).

1. Introduction

Federated Learning (FL) is a distributed machine learning framework that has emerged to address data privacy and security challenges in the field of artificial intelligence. With its decentralized data management approach and collaborative optimization mechanism, FL enables multiple parties to jointly train models and share knowledge without exposing raw data, thereby balancing model performance and data security. It provides a feasible pathway for cross-institutional and cross-domain data collaboration. In recent years, this technology has been widely applied in areas such as healthcare, financial risk control, intelligent transportation, and the Internet of Things (IoT), offering new solutions for multi-party cooperation and privacy-preserving computation. Based on differences in data features and sample distributions, federated learning can be categorized into three typical paradigms—horizontal, vertical, and transfer federated learning—each designed to address discrepancies in feature space or sample space. This diversified modeling mechanism not only enhances the generalization capability of models but also expands the application boundaries of artificial intelligence in compliant and secure environments. Therefore, federated learning is gradually becoming a key technology for building a trustworthy and sustainable intelligent ecosystem.

In recent years, with the rapid development of artificial intelligence, deep learning techniques have been widely applied across multiple domains. For example, in the medical domain, deep learning methods have been successfully applied to medical image analysis tasks, contributing to disease detection and diagnosis [1]. In the field of big data analytics, deep learning techniques have also been extensively employed to analyze and learn from large volumes of unstructured data, thereby supporting various data-intensive applications [2]. Moreover, deep learning has also been extensively studied and applied in agriculture. For instance, an intelligent agricultural system based on a hybrid CNN-SVM model enables the early detection of plant diseases, which significantly enhances agricultural productivity and provides effective safeguards [3]. It also significantly advancing

research in the field of biometrics. As one of the most representative models, Convolutional Neural Networks (CNN) have been extensively utilized for fingerprint recognition and classification tasks. For example, Militello conducted a systematic comparison of several CNN models, confirming their efficiency and accuracy in fingerprint recognition and classification [4]. Similarly, Al-Wajih demonstrated that CNN-based methods achieved promising performance in classifying fingerprint patterns such as arches, loops, and whorls [5]. More recently, Garg proposed an improved approach based on existing CNN architectures by incorporating data augmentation strategies and inversion methods, which significantly enhanced recognition accuracy under more complex acquisition conditions [6]. Although these CNN-based models have made remarkable progress in terms of recognition accuracy, issues concerning security and privacy during the recognition and classification process remain inadequately addressed. To mitigate such security challenges in fingerprint classification, federated learning has been introduced into biometric recognition research. As a distributed machine learning paradigm, federated learning enables multiple participants to perform model training by keeping raw data on local devices, while uploading the locally trained model parameters to the central server simultaneously to iteratively upgrade the central large model. The kind of architecture constructs a more comprehensive global model with improved efficiency and accuracy, effectively reducing the risk of privacy leakage. Guo further discussed how federated learning can be applied to biometric recognition field, emphasizing its great potential in enhancing recognition performance and safeguarding data security, particularly in privacy-sensitive applications such as fingerprint recognition and classification [7].

The aim of this study review is to analyze and study deep learning large models based on federated learning methods to solve the potential threats faced by traditional models in privacy data security. The remainder of this paper is organized as follows: The second section will focus on describing and analyzing the structural characteristics of various federated learning models, their practical advantages and disadvantages in applications, as well as solutions to data security issues. The third section will not only delve into some main challenges of federated learning methods but discuss its future development prospects. Finally, this paper will conclude with a discussion on how federated learning models can address privacy and data security concerns.

2. Method

2.1. Horizontal Federated Learning (HFL)

HFL, a type of federated learning partitioned by data samples, is characterized by federation among datasets with low sample overlap but high feature overlap [8]. It is suitable for scenarios where the feature spaces of different clients are similar, but their user groups differ. In other words, clients share similar data feature structures but serve distinct sets of users.

2.1.1 Model Workflow (FedAvg as an Example)

First, a central server establishes a foundational model framework. Then, the server distributes this model to all clients, who train it locally using their own data to obtain locally updated model parameters. Next, clients encrypt these updated parameters and transmit them back to the central server. Finally, the server collects all local updates and aggregates them using algorithms such as weighted averaging (e.g., FedAvg). Through iterative communication and aggregation cycles, the global model is continuously refined and updated to form the next generation of the model.

2.1.2 Advantages and Disadvantages

In terms of advantages, the architecture is relatively simple, and algorithms such as FedAvg are easy to implement and deploy due to their clear logic. HFL also offers high communication efficiency, since only model parameters are transmitted, resulting in smaller data transfer volumes. Moreover, it is widely applicable to scenarios where participants share similar data features but are geographically or organizationally distributed, such as mobile device networks and IoT systems.

However, HFL requires high data homogeneity, as all participants must have aligned feature spaces. It also faces statistical heterogeneity challenges, where non-independent and non-identically distributed (non-IID) local data can lead to convergence difficulties and reduced model performance. Furthermore, client participation bias may occur, since clients with differing levels of activity can cause uneven model updates.

2.1.3 Practical Application Scenarios

Horizontal federated learning is commonly applied in scenarios where the feature spaces are similar but user groups differ. For instance, two companies developing fingerprint recognition systems may possess datasets with similar fingerprint feature structures, but their users come from different countries or regions. If the two companies can adopt the model for collaboration, the model trained by them may enhance recognition accuracy easily without exchanging sensitive raw data.

In addition, HFL has been applied in a wide range of domains, including smart devices, Internet of Things (IoT), and healthcare. In recent years, researchers have also extended its applications to computer vision and image recognition tasks. For example, Mandal et al. demonstrated that horizontal federated learning models exhibit excellent efficiency and performance in object detection, recognition, and segmentation tasks [9].

2.1.4 Risks and Solutions

In terms of risks, in horizontal federated learning (HFL) models, since all of the clients use the same model architecture, attackers may reconstruct training samples or infer original data features through uploaded gradients or model parameters — a phenomenon known as the Model Inversion Attack. Therefore, the primary focus of security protection lies in preventing gradient information leakage and safeguarding the model aggregation process.

However, there are also some solutions to these risks. Firstly, Secure Multi-Party Computation (MPC): Each client participates in encrypted computation without revealing specific parameters, ensuring that the server can only obtain aggregated results. Common methods include Secret Sharing and Random Masking mechanisms. Secondly, Homomorphic Encryption (HE): It not only enables arithmetic operations on encrypted data but also supports model aggregation without decryption. Classical schemes include Paillier encryption and CKKS encryption (suitable for floating-point gradients). Thirdly, Differential Privacy (DP): Adds noise to the gradients before uploading to prevent the influence of individual samples from being detected. A representative algorithm is DP-FedAvg technique, which effectively balances privacy preservation with model performance.

2.2. Vertical Federated Learning (VFL)

VFL, a type of federated learning partitioned by data features, is characterized by collaboration among datasets with high sample overlaps but low feature overlap. It is suitable for scenarios where the sample spaces of different clients largely overlap, while their feature spaces differ. In other words, the clients share many of the same users, but each holds different types of information about them.

2.2.1 Model Workflow

First, the model must align samples across parties. Using privacy-preserving intersection protocols, it identifies the shared sample subset among participants without revealing any non-overlapping samples. Next, each participant performs encrypted local training on the aligned samples, computing model parameters such as gradients or losses associated with its own feature subset. These intermediate parameters are then exchanged and aggregated via secure protocols to update the global model collaboratively.

Throughout this process, no participant has access to another's raw data or complete training parameters. All of the participants only retain the part of the model corresponding to their own feature space. During inference, the participants collaborate again—using secure computation protocols—to jointly compute prediction results without exposing private data.

2.2.2 Advantages and Disadvantages

In terms of advantages, Vertical federated learning unlocks the value of distributed data by breaking down “data silos,” allowing different feature dimensions to be combined for enhanced model performance. It offers high commercial value, particularly for cross-industry or cross-company collaborations such as finance and healthcare. This advantage is well illustrated in the MMVFL framework proposed by Siwei Feng [10], which enables multiple data owners to leverage information from others to improve task performance without disclosing raw data.

However, technology involves high complexity, requiring extensive cryptographic operations (e.g., homomorphic encryption, secret sharing), which leads to significant computational and communication costs. Sample alignment introduces privacy risks—although privacy-preserving intersection techniques can protect non-overlapping samples, the protocol itself may still leak information. Additionally, VFL systems are architecturally complex, requiring multi-party coordination, making deployment and maintenance costly. As Folino et al. [11] pointed out, under the constraints of cost and privacy risk, designing a system that achieves both efficiency and strong privacy protection remains a major challenge.

2.2.3 Practical Application Scenarios

VFL is particularly suitable for cases where the sample space overlaps substantially, but feature spaces differ. For example, a hospital and an insurance company may share many of the same clients—the hospital holding medical data while the insurer holds financial and claims data. Using VFL, the two can collaboratively train a more accurate and efficient disease risk prediction model without sharing raw information. A recent study proposed a neural network-based VFL framework with built-in privacy-preserving mechanisms, demonstrating strong potential for collaborative applications in areas such as healthcare and financial risk prediction [12].

2.2.4 Risks and Solutions

In terms of risks, in vertical federated learning (VFL), model training involves encrypted computation and gradient exchange across different feature spaces, which poses greater security challenges. For instance, improper intersection computation during sample alignment can lead to privacy leakage at the alignment stage.

However, there are also some solutions about these risks. Firstly, Private Set Intersection (PSI): Used to identify the common sample set among participants. Common implementations include Elliptic Curve-based PSI (ECC-PSI) and Homomorphic Encryption-based PSI. Classic applications include Google’s and Tencent’s VFL systems. Secondly, Encrypted Gradient Exchange: Employs homomorphic encryption or secret sharing to protect gradient information, preventing the server or other clients from inferring original features. A well-known framework is SecureBoost, which uses encrypted gradient boosting trees for cross-institutional modeling. Thirdly, Hybrid Secure Aggregation with Differential Privacy: During gradient transmission, this approach combines noise perturbation with encryption to achieve dual-layer protection. For example, Alibaba’s Hybrid Secure VFL Framework integrates HE and DP for privacy-preserving medical data analysis.

2.3. Federated Transfer Learning (FTL)

FTL is designed for scenarios characterized by small overlap in both sample and feature spaces among participants. In such cases, traditional federated learning performs poorly because there is limited shared information between parties. Therefore, transfer learning techniques are introduced to facilitate knowledge transfer between related domains, compensating for the lack of shared data and improving overall model performance.

Li’s research indicates that this technique not only facilitates collaboration between heterogeneous participants, but also effectively addresses the dual challenges of data scarcity and privacy preservation, thereby expanding the applicability of federated learning to a broader range of scenarios [13].

2.3.1 Model Workflow

Federated transfer learning combines characteristics of both horizontal and vertical federated learning but places greater emphasis on discovering mapping relationships between different feature and sample spaces. Typically, the central server extracts a shared latent feature space from models trained locally by each participant. Then, using a small shared sample set or knowledge distillation, knowledge from one party is transferred to another. During the training phase, each participant learns domain-specific parameters locally while simultaneously contributing to the shared latent representation. Wei [14] proposed a domain-adaptive meta-learning framework that achieves effective knowledge transfer across non-overlapping feature spaces through secure aggregation and alignment of latent embeddings. Throughout this process, strict secure communication protocols are required to protect intermediate information exchanged during transfer.

2.3.2 Advantages and Disadvantages

In terms of advantages, FTL extends the application boundaries of federated learning, solving problems in data- and feature-scarce environments. It holds great potential for enabling cross-domain knowledge integration and innovation.

However, FTL is the least mature among the federated learning paradigms, with limited research and practical implementation. Its effectiveness is uncertain, as transfer learning may fail depending on the degree of domain mismatch. Moreover, it has high computational and communication overhead, combining the complexities of both horizontal and vertical approaches.

2.3.3 Practical Application Scenarios

Federated transfer learning is ideal for cases with small overlaps in both sample and feature spaces. For example, two e-commerce platforms operating in different countries with distinct user bases and product categories can collaboratively enhance their recommendation systems using FTL. In addition, Chen [15] proposed an FTL strategy for cross-device fault diagnosis in industrial systems. Their framework successfully transferred diagnostic knowledge among heterogeneous devices, demonstrating the effectiveness of FTL in Industrial Internet of Things (IIoT) environments.

2.3.4 Risks and Solutions

In terms of risks, Federated transfer learning (FTL) fundamentally involves cross-domain federated knowledge transfer. The major security risks include potential information leakage during knowledge distillation, and privacy inference attacks where adversaries exploit shared embedding representations to infer the other party's private data.

However, there are also some solutions to these risks. Firstly, Secure Knowledge Distillation: Applies perturbation or encryption to model outputs (e.g., logits) to prevent reverse inference of original feature distributions. Classical approaches include Federated Distillation (FD) and PrivateFD. Secondly, Secure Embedding Alignment: Each party performs encryption or noise injection in the shared latent feature space to ensure the safety of mapping functions. A representative study is Wei et al. [14], who proposed a Domain-Adaptive Meta FTL framework for secure cross-domain knowledge transfer. Thirdly, Multi-layer Privacy Protection Strategy: Combines differential privacy, MPC, and adversarial defense mechanisms to mitigate cross-domain information leakage risks. For instance, Chen et al. [15] introduced a privacy-enhanced cross-device FTL framework for industrial Internet of Things (IIoT) applications, demonstrating its effectiveness in secure collaborative diagnostics.

3. Discussion

3.1. Challenges

3.1.1 Statistical heterogeneity problem

The data distributions across clients are often different (in terms of feature distributions or data quality), causing local updates to deviate from the global optimization objective. This leads to unstable training, slow convergence, and degraded overall model performance. The global model may generalize poorly, and certain clients may experience significantly lower performance compared to their standalone local models, with their personalization needs unmet.

3.1.2 System and communication bottleneck problem

Federated learning (FL) relies on frequent communication between clients and the central server. Edge devices may have limited computational power or energy capacity, while unstable networks, offline nodes, or asynchronous updates can affect efficiency and consistency. Consequently, training becomes time-consuming, communication costs are high, and large-scale deployment is challenging.

3.1.3 Privacy and security attack problem

Even without sharing raw data, gradients, model updates, or intermediate parameters may still leak sensitive information. Malicious clients could launch data poisoning or backdoor attacks, leading to user privacy breaches, model hijacking, and reduced system trustworthiness.

3.1.4 Incentive mechanism and fairness problem

Participants contribute unequally (in terms of data volume, data quality, and computational resources), yet there is often no clear, trustworthy incentive or reward distribution mechanism. The model may also perform unfairly across different participant groups, which can result in low client engagement, unsustainable long-term collaboration, and ethical or regulatory risks.

3.1.5 Regulatory compliance and cross-domain interoperability problem

Cross-country or cross-organizational collaboration is restricted by regulations such as the GDPR, privacy laws, and data localization requirements. Additionally, differences in data semantics and schema across institutions hinder smooth integration. Achieving real-world industrial deployment is therefore complex and requires both legal and technical guarantees.

3.2. Future Prospects

3.2.1 Standardization and Governance Framework for Cross-Domain Interoperability

The emergence of preliminary industry standards—covering communication protocols, compliance auditing interfaces, and regulatory inspection mechanisms—supported by pilot projects recognized by industry or regulatory bodies.

3.2.2 Federated Multimodal and Cross-Modal Learning Becoming Mainstream

Enabling joint modeling of diverse data types such as text, images, and sensor data under privacy constraints (e.g., combining medical imaging, electronic health records, and genomics for integrated healthcare analysis).

3.2.3 Deep Integration of Federated Learning and Edge Intelligence (FEL)

Achieving collaborative learning and inference across the end–edge–cloud hierarchy (front-end devices, edge servers, cloud servers) in real-time systems such as intelligent transportation and smart grids.

3.2.4 Maturation of a Trustworthy Federated AI Ecosystem and Governance

Establishing comprehensive frameworks for regulation, auditing, incentive mechanisms, and cross-domain compliance, enabling large-scale deployment in national and industrial sectors such as healthcare, finance, and public safety.

4. Conclusion

This study conducts an analytical review of federated learning models to explore their approaches to addressing data security challenges. By examining the architectural characteristics, workflows, and advantages and disadvantages of various federated learning models, the paper identifies the most suitable real-world application scenarios for each model and, based on the analysis, proposes potential risks, challenges, and corresponding data security solutions. Through practical analysis, the results indicate that although federated learning models—with their decentralized data management and collaborative optimization mechanisms—facilitate secure collaboration across domains and institutions, they still face several pressing challenges today, such as regulatory compliance and cross-domain interoperability problems and privacy and security attack risks. The paper suggests mitigating these risks through strategies such as multi-layer privacy protection and encrypted gradient exchange. Finally, the essay envisions future developments in the field, including the emergence of federated multimodal and cross-modal learning as a mainstream paradigm.

References

- [1] Litjens M, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*. 2022.
- [2] Khan MA, et al. Big data analytics deep learning techniques and applications: A survey. *Information Systems*. 2024.
- [3] Kamilaris M, Prenafeta-Boldú FX. Smart agriculture applications using deep learning technologies: A survey. *Applied Sciences*. 2022; 12 (12): 5919.
- [4] Militello C, Rundo L, Battiato F, Conoci S. Fingerprint classification based on deep learning. *Symmetry*. 2021 May; 13 (5): 750.
- [5] Al-Wajih YA. Finger type classification with deep convolution neural networks. In: *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. 2022. p. 611–618.
- [6] Garg R, Aggarwal R, Nain S, Agarwal V. Fingerprint recognition using convolution neural network and image processing. *Journal of Information Security and Applications*. 2024; 76: 103597.
- [7] Guo J, Mu H, Liu X, Han C, Sun Z. Federated learning for biometric recognition: A survey. *Artificial Intelligence Review*. 2024; 1–36.
- [8] Khitrov I, Stenger M, Gamer T, Kellerer W. Accuracy tradeoffs of federated learning approaches. In: *Proceedings of the ITG Specialist Conference on Network Architectures and Services (NET)*. Munich, Germany; 2022. p. 1–6.
- [9] Mandal PK, Kim K, Lee SJ. Horizontal federated computer vision. *ACM Transactions on Reproducible Computing (ACM TRC)*. 2024; 17 (2): 1–18.
- [10] Feng S, Chen Y, Lin W. MMVFL: A simple vertical federated learning framework. *Frontiers in Computational Neuroscience*. 2024; 18: 10820651.
- [11] Folino F, Folino G, Pisani FS, Pontieri L. Efficiently approaching vertical federated learning by combining data reduction and conditional computation techniques. *Journal of Big Data*. 2024; 11 (45): 1–19.
- [12] Anees A, Field M, Holloway L. A neural network-based vertical federated learning framework with server integration. *Engineering Applications of Artificial Intelligence*. 2024 Dec 1; 138: 109276.
- [13] Li X, Zhang L, Chen Y, Wang J. A comprehensive survey of federated transfer learning. *Frontiers of Computer Science*. 2024; 18 (4): 184006.
- [14] Wei H, Zhang X, Liu Y. A federated transfer learning framework based on domain adaptation for student classification. *Applied Sciences*. 2022; 12 (21): 10711.
- [15] Chen Y, Li F, Zhou T. Federated transfer learning strategy: A novel cross-device fault diagnosis framework. *Sensors*. 2023; 23 (15): 6721.