10.48047/jocaaa.2025.34.12.02

Federated Learning: Collaborative Machine Learning Without Data Sharing

Sudhakar Kandhikonda
Birla Institute of Technology and Science, Pilani (BITS Pilani), India

Abstract

Federated learning is a radically new model of machine learning methods, which allows the development of models on top of cooperative distributed devices or institutions without the need to centralize raw data. This new method ensures privacy and sovereignty and enables participants to enjoy the collective intelligence, which is a significant issue in privacy-related fields. It operates as an iterative process, where a central server synchronizes model training updates from participants' local computing and communicates only model parameters, not the original data. This article looks at the basic architecture of federated learning systems, communications efficiency protocols, approaches to statistical and system heterogeneity, and improved privacy protocols like differential privacy, secure aggregation, and homomorphic encryption. Federated learning overcomes traditional obstacles to cooperation in controlled settings, as seen in applications across healthcare, finance, mobile computing, and smart city infrastructure. Although such benefits can be achieved, issues such as the vulnerability to model poisoning, the possibility of inversion attacks, and the computing limitations of edge devices should be addressed through continuous research. With the heightened privacy issues and regulatory demands across the world, federated learning presents an excellent model of responsible AI development that fulfills the innovation and data protection needs.

Keywords: Privacy-Preserving Machine Learning, Decentralized Model Training, Secure Multi-Party Computation, Data Sovereignty, Edge Computing Intelligence

1. Introduction

Traditional machine learning methods typically require the centralization of data from various sources into a single repository for model training. Such consolidated architecture introduces significant vulnerabilities in dealing with sensitive information, such as personal identifiers, health records, or financial transactions. McMahan et al. introduced federated learning as a decentralized approach to machine learning that addresses these concerns of privacy while enabling collaborative model training [1]. Their work demonstrated how distributed devices could collectively train shared prediction models while keeping all training data local, fundamentally changing the traditional paradigm of data collection and model development.

The concept of Federated Learning is a transformation in paradigm that implies model training with no direct exchange of raw data among parties. Such a methodology is implemented to solve major issues in actual-world scenarios where bandwidth constraints, intermittent connectivity, and privacy concerns render centralized learning impractical. A central server coordinates the process of training in a federated system, but computation occurs on client devices using their local data. Only the model updates, not the raw data itself, are transferred to the server, significantly reducing the risk associated with data privacy. McMahan et al. showed this very approach with the Federated Averaging algorithm, which aggregates these locally computed updates into a single global model without touching the underlying training data [1].

10.48047/jocaaa.2025.34.12.02

The technical architecture of FL embodies key principles that set it apart from centralized approaches: the data stays distributed across clients, the selection of participants is based on availability criteria, and communication efficiency becomes paramount given network constraints. These principles manifest in specialized optimization algorithms designed for federated settings, where non-IID data distributions across clients present unique challenges compared to traditional machine learning assumptions.

Privacy preservation extends beyond keeping data localized. As Bonawitz et al. discuss, even model updates can potentially leak information about the underlying training data [2]. Their work introduces cryptographic techniques that enable a server to compute the sum of model updates from multiple users without learning individual contributions. This secure aggregation protocol ensures the server only sees the final aggregated update after sufficient users have contributed, providing formal privacy guarantees even against curious servers while addressing practical challenges like dropped connections and device failures.

Federated learning has lately been applied beyond mobile devices in vertical industries such as health, finance, and other sensitive domains where privacy issues have so far held back advanced machine learning. This opens opportunities for multi-organizational collaboration that was previously impossible because of competitive, regulatory, or ethical reasons and hence demands systems that are designed to operate at scale across heterogeneous devices with various capabilities and availability patterns [2].

2. Fundamentals of Federated Learning

Federated learning works in a decentralized model wherein the training of the model happens locally across different devices or institutional silos, rather than in a centralized data repository. This kind of paradigm shift fundamentally changes how machine learning systems are developed in privacy-sensitive domains. The process can be conceptualized to be somewhat like a group of master chefs jointly perfecting a recipe through discussion of refinements and modifications without the need to disclose their secret ingredients or proprietary techniques. Each chef (device or institution) experiments with the recipe using his own unique ingredients (local data) and then shares his procedural refinements (model updates), not the specific constituents that form his secret ingredients. This, in turn, results in a superior final product without losing any intellectual property or privacy of each contributor.

The federated learning workflow is designed in a systematic iterative process that keeps data localized, hence enabling collaborative intelligence. First, a centralized server initializes a global model and distributes it to participating nodes through secure communication channels. This initial model may be randomly initialized or pre-trained on publicly available data as described by Konečný et al. [3]. Second, each participating node independently trains the model on its local data using standard optimization techniques like stochastic gradient descent, adapting the global parameters to better fit local patterns and distributions. The training is fully on-device, and there is no external access to the raw training examples, which will ensure full data sovereignty during the training lifecycle.

Once the local training is complete, just model updates, generally in the form of gradients or parameter differences, are returned to the central server. This represents the critical privacy-preserving component of federated learning because the transmitted information contains derived knowledge, not raw data. Then, the central server uses sophisticated aggregation algorithms to piece together these heterogeneous updates into a coherent global improvement. According to the work of Yang et al., the Federated Averaging algorithm performs an average of the local updates weighted in proportion to the amount of training data at each node [4]. The described aggregation mechanism resolves the problem of non-IID data distribution, which naturally occurs when data is divided across different users or organizations.

10.48047/jocaaa.2025.34.12.02

The refined global model is then redistributed to the participating nodes for further training in the next round, starting yet another cycle of local refinement. This goes on in an iterative manner until the global model converges to some optimal state or reaches a pre-specified performance threshold. These stopping criteria can be related to plateaus in validation accuracy or loss function stabilization and may involve a maximum number of communication rounds. During this entire process, raw training data does not leave its original location; therefore, privacy remains ensured while the knowledge embedded across the distributed dataset is utilized effectively.

Stage	Process	Data Location	Privacy Level	Communication Load
1	Global Model Initialization	Central Server	High	Low
2	Model Distribution	Nodes/Devices	High	Medium
3	Local Training	Nodes/Devices	Maximum	None
4	Model Update Transmission	Central Server	High	Medium
5	Update Aggregation (FedAvg)	Central Server	High	Low
6	Improved Model Distribution	Nodes/Devices	High	Medium
7	Convergence Evaluation	Central Server	High	Low

Table 1: Federated Learning: Data Privacy vs. Communication Load Across Workflow Stages [3, 4]

3. Technical Implementation

Federated learning systems create several technical challenges for their implementation that do not exist in traditional centralized machine learning paradigms. These include communication constraints, data distribution variations, and system capability differences across the federation.

3.1 Communication Efficiency

The bandwidth limitation is a big bottleneck in the deployment of federated learning, especially in mobile and edge computing environments. The efficient communication protocols between the central server and edge devices are extremely important for practical implementation. Li et al. developed several seminal techniques for reducing communication overhead: structured updates, where the model changes are constrained to a lower-dimensional space, and sketched updates that compress the model changes using quantization, random rotations, and subsampling [5]. Their experiments showed that communication costs can be reduced by orders of magnitude with negligible effects on accuracy. Other approaches include gradient compression, where only the most significant gradient values are transmitted; quantization, which reduces the precision of the gradient representation; and sparsification, which zero-outs the less important updates selectively. All these tricks render federated learning feasible in bandwidth-restricted settings wherein conventional distributed learning would be infeasible.

3.2 Statistical Heterogeneity

Data between devices or between institutions tends to be non-IID, and the resulting heterogeneity can be very problematic in terms of model convergence and model performance.

This heterogeneity naturally occurs due to variations in user behavior, geographic differences, and institutional specializations. For instance, different hospitals will likely view vastly different patient demographics and condition prevalence in healthcare contexts. Li et al. suggest that this statistical heterogeneity manifests through two forms: feature distribution skew, where different input distributions exist across nodes, and label distribution skew, where there is variation in the prevalence of different

10.48047/jocaaa.2025.34.12.02

output classes [5]. Several advanced aggregation algorithms have been designed to handle these challenges. FedAvg adopts weighted averaging according to the amount of local data; FedProx adds a proximal term to prevent local updates from drifting too far from the global model; while SCAFFOLD uses variance reduction techniques to correct for client drift during training. Karimireddy et al. show that these methods significantly improve the convergence rate and final model performance compared to naive distributed optimization in heterogeneous settings [6].

3.3 System Heterogeneity

Participating devices or institutions generally have different computational capabilities, storage capacity, and network connectivity characteristics. This system heterogeneity adds more complexity to the deployment of federated learning. For example, mobile devices may have limited batteries and processing power, while edge servers may have occasional connectivity but with higher computational resources. As pointed out by Karimireddy et al., naively ignoring these differences leads to inefficient resource utilization and biased models that favor the contributions from more capable devices [6]. Consequently, several adaptive algorithms have been developed to handle such heterogeneity, including asynchronous aggregation protocols that allow not all devices to finish their training simultaneously; selective participation of devices that take into consideration current device status; and heterogeneous model assignments that allow model complexity to be matched to device capabilities. Optimizing resource utilization across the federation, these techniques enable high-quality models, thus truly allowing inclusive federated learning across diverse hardware ecosystems.

10.48047/jocaaa.2025.34.12.02

Challenge Type	Key Issue	Solution Approach	Effectiveness	Implementation Complexity
Communication Efficiency	Bandwidth Limitations	Structured Updates	High	Medium
	Data Transfer Volume	Sketched Updates	High	Medium
	Transmission Costs	Gradient Compression	Medium	Low
	Precision Requirements	Quantization	Medium	Low
	Update Size	Sparsification	High	Medium
Statistical Heterogeneity	Feature Distribution Skew	FedAvg Algorithm	Medium	Low
	Label Distribution Skew	FedProx Algorithm	High	Medium
	Client Drift	SCAFFOLD Method	High	High
System Heterogeneity	Varying Computational Power	Asynchronous Aggregation	Medium	High
	Device Availability	Selective Participation	High	Medium
	Resource Limitations	Heterogeneous Model Assignment	High	High

Table 2: Effectiveness vs. Implementation Complexity of Federated Learning Solutions [5, 6]

4. Enhanced privacy and security

While federated learning itself increases privacy by maintaining the raw data locally, an intelligent adversary can still manage to extract sensitive information through updates in models or inference attacks. Additional cryptographic and privacy-preserving techniques were developed for better protection of data in federated learning systems.

4.1 Differential Privacy

Differential privacy offers mathematical assurances regarding the extent to which a particular training example affects the output model so that information is not leaked. Differential privacy protocols do not allow the sharing of model parameters to be used to extract one of the individual training examples. Wei et al. formalize this approach through the introduction of client-level differential privacy, where the objective is to make the inclusion or exclusion of any client's entire dataset have a limited impact on the output distribution of the federated learning algorithm 7. The amount of privacy provided is usually

10.48047/jocaaa.2025.34.12.02

bounded by a so-called privacy budget ε, which is an upper limit of the information leakage about any individual's data. Smaller values of ε indicate higher privacy but at a reduced model utility. More recent works involve adaptive noise mechanisms that automatically adapt the noise distribution depending on the update characteristics to optimize the privacyutility tradeoff. It effectively mitigates the membership inference attacks in which an adversary aims to infer whether certain data points were used during training.

4.2 Secure Aggregation

Secure aggregation uses cryptographic protocols that allow computation on aggregate statistics without exposing any information about individual contributions to any of the parties, including the central server. As such, this solution would also offer robust protection against curious servers analyzing individual updates. Truex et al. leverage threshold homomorphic encryption for secure aggregation, whereby each client encrypts their model update with a collective public key; the server can then decrypt only the aggregated final result once a sufficient number of clients contribute their partial decryption keys [8]. By doing so, it protects individual updates even when the server colludes with a portion of the clients, ensuring that unless a threshold number of participants' keys are compromised, updates cannot be inferred. Additional security features include participant verification mechanisms to prevent sybil attacks, secure key exchange protocols, and dropout resilience to maintain operability during client disconnections throughout the aggregation process. This provides a secure computing environment for privacy-sensitive applications in federated learning deployed across domains such as healthcare and finance.

4.3 Homomorphic Encryption

Homomorphic encryption represents an advanced cryptographic methodology that enables computation directly on encrypted data without requiring decryption. This concept allows model training and inference while maintaining the privacy end-to-end, since data remains encrypted throughout the machine learning pipeline. Truex et al. illustrate that partially homomorphic encryption schemes, supporting specific operations such as addition or multiplication, can be applied efficiently to federated learning with acceptable computational overhead in [8]. These can therefore enable the server to carry out aggregation operations on encrypted model updates, such that the contribution of each of them remains confidential. Fully homomorphic encryption enables arbitrary computations on encrypted data, with stronger guarantees, and at the cost of massive computational of resource-constrained devices. The optimization of encryption schemes for machine learning tasks is still being studied, and specialized cryptosystems already trade off the security needs with the real performance limits. Such works gradually make homomorphic encryption more viable for the use of production federated learning systems, at least in those highly regulated industries where privacy considerations are paramount.

10.48047/jocaaa.2025.34.12.02

Privacy Technique	Security Level	Computational Cost	Privacy-Utility Tradeoff	Implementation Complexity	Threat Protection
Base Federated Learning	Medium	Low	Low	Low	Data Exposure
Differential Privacy	High	Medium	High	Medium	Membership Inference
Secure Aggregation	Very High	High	Medium	High	Curious Servers
Partial Homomorphic Encryption	High	High	Medium	High	Individual Updates
Full Homomorphic Encryption	Maximum	Very High	High	Very High	Complete Data Protection

Table 3: Comparative Analysis of Privacy-Utility Tradeoffs in Federated Learning Security Techniques [7, 8]

5. Applications in Privacy-Sensitive Domains

Federated learning has emerged as a transformational approach in domains where data privacy concerns, regulatory requirements, and competitive considerations have historically limited the application of advanced machine learning techniques. Allowing for collaborative model training without actual sharing of data, federated learning opens up new opportunities in many areas.

5.1 Health Care

Federated learning for medical applications involves the collaboration of healthcare research across institutions without violating patient confidentiality and regulatory requirements such as HIPAA, GDPR for healthcare contexts, and other regional data protection frameworks. Rieke et al. conducted pioneering work to examine the future of digital health with federated learning, emphasizing how this approach can address the problem of the fragmentation of medical data across institutions while maintaining privacy [9]. Their work articulates how federated approaches can surmount data silos in healthcare-that is, models can be trained across distributed datasets while sensitive patient information remains secure within each of their respective institutions. The authors stress several promising applications: medical imaging analytics, where patient scans can stay safely in their respective hospitals yet contribute to better models; clinical predictive modeling based on electronic health records; and personalized treatment recommendation systems. They point out that among the most critical issues in AI-implementation in the healthcare-data-accessibility domain, they include integration of heterogeneous sources and regulatory compliance, which can be addressed with the help of FL. This can especially be helpful in environments where sharing is limited by any legal or institutional restrictions or patient consent restrictions, in which smaller hospitals that may have smaller datasets can, in turn, benefit and contribute to models that have been trained on a diverse patient population.

5.2 Finance

10.48047/jocaaa.2025.34.12.02

Banking is a highly regulated sector, and confidentiality of data is a top priority in financial institutions. Federated learning provides the platform for these organizations to harvest collective intelligence without leaking any sensitive client information. Zheng et al. show how federated learning can be effectively applied to credit risk assessment, arguably one of the most mission-critical and privacy-sensitive tasks in financial services [10]. Their work explores ways financial institutions can develop better credit scoring models by collaboratively training across distributed data sources without revealing confidential client information and proprietary underwriting criteria. The paper develops a new federated learning framework for credit risk assessment that takes into account data heterogeneity across a variety of financial institutions while guaranteeing privacy and regulatory compliance. Their methods handle the class imbalance inherent in a credit default data set and attain superior results from using isolated institutional data. Other applications in the financial sector include fraud detection systems to identify suspicious patterns across institutions, anti-money laundering frameworks, which may benefit from enhanced detection with no sharing of transaction details, and market risk assessment models, which now benefit from better visibility into the market with no disclosure of trading strategies. These various applications allow financial institutions to leverage the advantages of collaborative intelligence while maintaining competitive separation and regulatory compliance.

5.3 Mobile Devices

Federated learning has been adopted very early by smartphone manufacturers and application developers to enhance personalized services while promising not to infringe on users' privacy. It has become particularly valuable for keyboard prediction models, voice recognition systems, and content recommendation algorithms, which get improved based on individual user preferences. It allows these systems to continuously improve without sacrificing either privacy preferences or network bandwidth requirements since data processing is local and only model updates are shared. Implementations of mobile federated learning address the challenge of intermittent connectivity through asynchronous update mechanisms that queue model changes until the network conditions are favorable. This methodology allowed making substantial improvements in the quality of user experiences while meeting increasingly strict privacy regulations and user expectations related to data handling.

5.4 Smart Cities

The peculiarities of managing urban infrastructure imply the appearance of specific privacy issues because large cities gather more and more specific data regarding transportation patterns, capital consumption, and the activities of citizens. If the municipal systems and departments are not required to centralize potentially sensitive information, federated learning will enable them to cooperate in solving optimization problems. They may be used in prediction models of traffic flow, which may tap into data of multiple transportation agencies; in optimisation of energy consumption systems, which may need to coordinate between utility providers; and in the field of public safety applications, which may need to respect jurisdictional boundaries whilst improving emergency response facilities. This solution is consistent with the principles of the responsible development of smart cities, where the advantages of data-driven governance are counterbalanced by the ideas of privacy and local control.

10.48047/jocaaa.2025.34.12.02

Domain	Application	Privacy Concern Level	Regulatory Complexity	Implementation Maturity
Healthcare	Medical Imaging Analysis	Very High	Very High	Medium
	Clinical Predictive Modeling	Very High	Very High	Low
	Personalized Treatment Systems	Very High	Very High	Low
Finance	Credit Risk Assessment	High	High	Medium
	Fraud Detection	High	High	High
	Anti-Money Laundering	Very High	Very High	Medium
	Market Risk Assessment	High	Medium	Medium
Mobile Devices	Keyboard Prediction	Medium	Medium	Very High
	Voice Recognition	High	Medium	High
	Content Recommendation	Medium	Medium	High
Smart Cities	Traffic Flow Prediction	Medium	Medium	Low
	Energy Optimization	Medium	Medium	Low
	Emergency Response Systems	High	High	Low

Table 4: Federated Learning Implementation Across Industry Domains: Privacy Concerns vs. Maturity [9, 10]

6. Challenges and Future Directions

Despite these promising advantages, federated learning is afflicted with some critical challenges, for which ongoing research will be needed to fulfill its promise for privacy-sensitive domains.

6.1 Model Poisoning

The distributed nature of federated learning introduces unique security vulnerabilities, particularly model poisoning attacks, in which malicious participants inject adversarial updates to compromise the performance or integrity of the global model. Detection may be most challenging in heterogeneous federated settings where non-malicious statistical outliers naturally occur. Fang et al. show that a small percentage of compromised clients (5-10%) may significantly degrade model accuracy or induce targeted misclassification when conventional aggregation methods are used [11]. In this study, several attack vectors are investigated, including label-flipping and gradient manipulation methods that can easily evade basic defenses. Advanced Byzantine-robust aggregation algorithms have also been developed, which can identify and filter possibly malicious contributions before incorporating them into the global model. These include the Krum, Bulyan, and Trimmed Mean approaches. Most of these defenses do not require any knowledge of underlying data distributions but usually depend on the statistical analysis of update

10.48047/jocaaa.2025.34.12.02

distributions to find outlying contributions. Reputation systems, which manage the quality of client contributions over time, complement these approaches with verifiable updates using secure multi-party computation protocols.

6.2 Model Inversion Attacks

In the case where raw data remains local, advanced attackers can still attempt to reconstruct training examples with model inversion attacks with shared model parameters. Such privacy violations are of particular concern in sensitive areas such as the health sector and finances.

Sun et al. have shown that gradient updates in standard federated learning implementations can leak significant information about the underlying training data, allowing adversaries to reconstruct input features with worrying fidelity [12]. Their work reveals that some neural network architectures are particularly vulnerable to such attacks during their early training phases. Developing stronger privacy-preserving techniques is an active research area; approaches range from differential privacy mechanisms adding calibrated noise to model updates, over secure aggregation protocols that block access to individual contributions, to knowledge distillation methods sharing only model predictions rather than parameter updates. Such means of protection inevitably bring about tradeoffs between privacy guarantees and model utility that must be carefully balanced depending on application needs, and face very high limits of computation, memory, and energy consumption. Specialized model architectures and training methodologies crafted for federated environments are required to overcome these limitations without sacrificing model quality. Other emerging approaches are model compression, quantization, and pruning, being adapted to a distributed context; split learning neural networks, which distribute components of a model across client computing devices and servers; and adaptive participation, which allocates computational demands to device capabilities. Such efficiency gains are needed to scale federated learning by high-capability devices to the level of democratizing privacy-preserving machine learning.

Conclusion

Federated learning is a radical paradigm of machine learning that essentially shapes the pursuit of technological progression alongside the doctrine of privacy protection. This methodology provides a chance to form cross-organizational and cross-device collaboration in areas that were difficult in the past due to privacy issues, competition, and legal mandates, because it allows collaborative development of models without the exchange of raw data. A federated approach solves some of the most important challenges of contemporary AI deployment, such as finding data silos, privacy concerns, and regulatory requirements, and it does not significantly harm model performance as compared with a centralized approach. With the growing power of computational applications on the edge and the growing importance of privacy issues across all sectors, federated learning will become the most common model in the responsible development of AI in sensitive areas. The current research on handling existing shortcomings in security, privacy assurances, as well as computational efficiency is likely to hasten adoption in the fields of healthcare, financial services, consumer technology, and infrastructure of the population. This privacy-preserving but collaborative approach is finally not only a technical breakthrough but also a paradigm shift to a more ethically minded artificial intelligence that does not violate individual privacy but utilizes collective intelligence.

10.48047/jocaaa.2025.34.12.02

References

[1] H. Brendan McMahan et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data," arXiv:1602.05629, 2023. https://arxiv.org/abs/1602.05629

[2] Keith Bonawitz et al., "Towards Federated Learning at Scale: System Design," arXiv:1902.01046v2, 2019. https://arxiv.org/pdf/1902.01046

[3] Jakub Konečný et al., "Federated Learning: Strategies for Improving Communication Efficiency," arXiv:1610.05492, 2017. https://arxiv.org/abs/1610.05492

[4] Qiang Yang et al., "Federated Machine Learning: Concept and Applications," arXiv:1902.04885v1, 2019. https://arxiv.org/pdf/1902.04885

[5] Tian Li et al., "Federated Learning: Challenges, Methods, and Future Directions," arXiv:1908.07873v1, 2019. https://arxiv.org/abs/1908.07873

[6] Sai Praneeth Karimireddy et al., "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning," arXiv:1910.06378v4, 2021. https://arxiv.org/pdf/1910.06378

[7] Kang Wei et al., "Federated Learning with Differential Privacy: Algorithms and Performance Analysis," arXiv:1911.00222, 2019. https://arxiv.org/abs/1911.00222

[8] Sai Praneeth Karimireddy et al., "SCAFFOLD: Stochastic Controlled Averaging for Federated Learning," arXiv:1910.06378v4, 2021.https://arxiv.org/pdf/1910.06378

[9] Nicola Rieke et al., "The Future of Digital Health with Federated Learning," arXiv:2003.08119v2, 2021. https://arxiv.org/pdf/2003.08119

[10] Chul Min Lee et al., "Federated Learning for Credit Risk Assessment," ResearchGate, 2023. https://www.researchgate.net/publication/364166166_Federated_Learning_for_Credit_Risk_Assessment

[11] Minghong Fang et al., "Local Model Poisoning Attacks to Byzantine-Robust Federated Learning,". https://www.usenix.org/system/files/sec20summer_fang_prepub.pdf

[12] Ziteng Sun et al., "Can You Really Backdoor Federated Learning?" arXiv:1911.07963, 2019. https://arxiv.org/abs/1911.07963