Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Federated Learning in Cloud Computing: A Novel Approach to Decentralized Data Processing and Privacy Preservation

Deepali Shukla[1], Dr. Kavi Bhushan[2]
[1]*Scholar M.Tech, Department of Computer Science & Engineering, Sir Chhotu Ram Institute of Engineering and Technology, Chaudhary Charan Singh University, Meerut, U.P., India*
[2]*Assistant Professor Sir Chhotu Ram Institute of Engineering and Technology, Chaudhary Charan Singh University, Meerut, U.P., India*

ARTICLE INFO

Received: 18 Dec 2024
Revised: 10 Feb 2025
Accepted: 28 Feb 2025

ABSTRACT

Federated Learning (FL) is a decentralized machine learning platform that allows for cooperative model training across several devices while maintaining data privacy and security. This paper explores the integration of FL into cloud computing environments, offering a novel solution to data centralization issues that raise privacy concerns. Cloud computing, with its centralized data processing model, often exposes sensitive information to potential breaches. In comparison, federated learning allows for the creation of machine learning models on decentralized devices without having to transfer sensitive data to central servers, thus ensuring better privacy preservation. The study outlines the proposed framework of FL within cloud systems, focusing on maintaining data confidentiality and optimizing computational efficiency. We assess key algorithms like Federated Averaging (FedAvg) and their performance in cloud-based scenarios. Experimental results demonstrate that FL can reduce communication overhead, achieve comparable model accuracy, and effectively enhance privacy in distributed settings. By evaluating multiple client configurations and using datasets like MNIST and CIFAR-10,The findings suggest that Federated Learning maintains privacy while facilitating scalable, efficient, and decentralized data processing within cloud settings. This study adds to the expanding knowledge surrounding Federated Learning, highlighting its potential for wide-scale deployment in industries where data privacy is paramount, such as healthcare, finance, and smart infrastructure. Additionally, the study explores the future possibilities for improving FL algorithms, considering advancements in edge computing, federated transfer learning, and adaptive learning models.

Keywords: Federated Learning, Cloud Computing, Data Privacy, Decentralized Data Processing, Federated Averaging, Model Aggregation, Privacy Preservation, Distributed Learning, Edge Computing, Communication Efficiency, Resource Optimization.

1. Introduction

1.1 Background

Cloud computing has fundamentally transformed data processing and storage paradigms by providing scalable, on-demand access to computing resources. It enables organizations to leverage vast computational power without the need for significant infrastructure investments. However, despite its many advantages, the centralization of data within cloud environments introduces considerable privacy and security risks, especially in industries that manage sensitive data, such as healthcare, banking, and the Internet of Things (IoT). Traditional cloud computing architectures transport data from end-user devices to centralized servers for processing, which increases the likelihood of data breaches and complicates adherence to data privacy laws such as GDPR and HIPAA [6][16]. This centralized approach not only heightens the risk of exposure but also undermines the privacy of individuals, posing a significant challenge in the era of data-centric applications.

1.2 Problem Statement

The centralization of data storage in cloud computing environments significantly compromises user privacy and data security. With more and more organizations transitioning to cloud-based solutions, infrastructures, the necessity for

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

approaches that facilitate secure and privacy-preserving collaborative learning becomes more critical. Federated Learning (FL) presents a promising approach to this issue by allowing for model training in a decentralized manner, eliminating the necessity of transferring raw data to central sites. While FL has been explored in various domains, its integration within cloud computing environments remains an under-explored area, necessitating further investigation into its potential for improving privacy and security.

1.3 Objective

This study's primary purpose is to integrate Federated Learning within cloud computing systems to enhance privacy, optimize communication efficiency, and maintain model performance. This study aims to analyze the effectiveness of Federated Learning algorithms in simulated cloud environments, focusing on their ability to improve data privacy, reduce communication overhead, and optimize resource utilization. By comparing the performance of FL models with traditional centralized learning models, the study offers an understanding of the advantages and obstacles associated with implementing Federated Learning in real-world cloud applications.

2. Literature Review

Federated Learning (FL) is a distributed machine learning paradigm that allows models to be trained across several decentralized devices or servers that have local data samples without having to share them. This method enhances privacy by keeping the original data on the client device and only exchanging model changes (such as gradients or weights) with a central server. for the purpose of aggregation. The fundamental components of FL include local model training, a secure aggregation process (e.g., using federated averaging), and efficient communication protocols to manage updates between clients and servers [4].

2.1 Current Research

Federated Learning has found significant applications across several data-sensitive domains:

- Healthcare: FL is increasingly used to train models across hospitals and medical institutions without directly accessing sensitive patient data. For example, Yang et al. showed that federated learning is applicable for developing models that predict patient outcomes while adhering under the requirements established by the Health Insurance Portability and Accountability Act (HIPAA) [5].
- Finance: In the financial sector, FL has been applied to develop fraud detection and credit scoring models while preserving customer data privacy. Zhao et al. [10] introduced an FL-based approach where financial institutions jointly develop machine learning models without moving user data to centralized servers.
- Internet of Things (IoT): Federated Learning has enabled smart devices, such as home assistants and wearable gadgets, to learn from local user data while minimizing cloud dependence. Li et al. [7] presented a privacy-preserving FL framework for smart homes that enables real-time personalization without compromising user privacy.
- Edge Computing: The merger of federated learning with edge computing has resulted in decreased latency and enhanced resource management. Chen et al. [4] showcased a federated learning framework designed for edge networks, showing performance gains in latency and bandwidth utilization.
- Methods for Protecting Privacy: Techniques Techniques like Differential Privacy and Secure Multi-party Computation (SMC) have been integrated with federated learning (FL) to improve data security. Abadi et al.[2] proposed a stochastic gradient descent algorithm that ensures differential privacy which can be incorporated into FL frameworks to ensure that no single data point contributes disproportionately to the final model.

2.2 Gap Analysis

Despite significant progress in applying Federated Learning to domains such as healthcare, finance, and IoT, its integration within large-scale cloud computing infrastructures remains relatively underexplored. Current research has primarily emphasized mobile and IoT environments [5][7], often overlooking critical cloud-native requirements such as elastic scalability, multi-tenancy, and distributed orchestration. Kairouz et al. [6] identified these limitations

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

as key open challenges in Federated Learning, stressing the importance of developing frameworks that can effectively merge the decentralized benefits of FL with the computational power and scalability of cloud platforms.

3. Research Methodology

The suggested structure leverages a cloud-based federated learning architecture, where a centralized cloud server coordinates multiple distributed client devices. Each client possesses its own local dataset and performs training independently, guaranteeing that unprocessed data remains on the local device. This configuration minimizes privacy threats while leveraging the processing power of edge devices.

The cloud server acts as the orchestrator, receiving periodic model updates (rather than raw data) from clients and performing global model consolidation. This method is continued throughout several communication cycles until the global model converges.

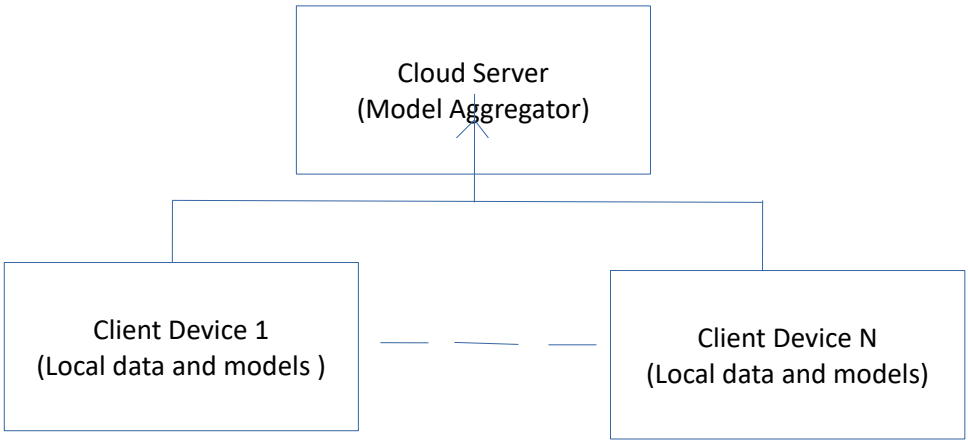

Figure 3.1: Proposed Framework for Federated Learning in Cloud Computing

3.1 Data Distribution Strategy

In the proposed system,training data is not IID (not identically independently distributed). and inherently heterogeneous across clients. Each participating client device holds a distinct portion of data (e.g., based on user activity, region, or sensor type). By design, no raw data is transmitted to a cloud server. Instead, after instructing the locals model on each client's dataset, only the modified [11]. This data partitioning ensures:

- Preservation of user privacy and regulatory compliance (e.g., GDPR, HIPAA).
- Decentralized training with minimal data movement, optimizing bandwidth use.

3.2 Algorithm Selection: Federated Averaging (FedAvg)

The Federated Averaging (FedAvg) algorithm is employed as the core mechanism for collaborative model training in this framework. FedAvg is particularly effective in scenarios characterized by limited communication bandwidth and heterogeneous (non-IID) data distribution, making it highly suitable for cloud-based federated systems. The algorithm combines local model updates from distributed clients through weighted averaging, thereby reducing communication costs while maintaining strong model performance [8]. It works as follows:

FedAvg Algorithm Steps

1. Initialization
   The server sets up a universal model with parameters wo.
2. Selecting Clients (Round t)

   Each round, a random subset $C_t \subseteq K$ is chosen from all clients.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. Client Training at the Local Level
   Each chosen client $k \in C_t$ trains the model on its local data D_k using stochastic gradient descent (SGD) for a few epochs:

$$w_k^{t+1} = w_k^t - \eta \nabla L(w_k^t, D_k)$$

   where η represents the learning rate, while L denotes the local loss function [12].

4. Model Update Transmission
   Clients transmit their revised weights w_k^{t+1} to the cloud server.

5. Global Model Aggregation
   The server performs weighted averaging over all participating clients' updates:

$$w^{t+1} = \frac{1}{|C_t|} \sum_{k \in C_t} w_k^{t+1}$$

6. Iteration and Convergence
   Steps 2 through 5 are reiterated until the global model reaches convergence to an optimal performance threshold.

3.3 Process Flow of Federated Learning

The process flow of Federated Learning illustrates how the interaction between the cloud server and distributed clients takes place in an iterative manner to collaboratively build a global model. This cycle is designed to ensure data privacy by keeping sensitive information on local devices while still allowing collective intelligence through secure model aggregation. By leveraging multiple communication rounds, the system balances computational efficiency with convergence accuracy, making it suitable for large-scale distributed environments.

The flowchart includes the following stages:

- Server initializes the global model.
- A subset of clients is selected.
- Clients perform local training and send updates.
- Server aggregates updates and distributes the new model.
- Repeat until model convergence.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

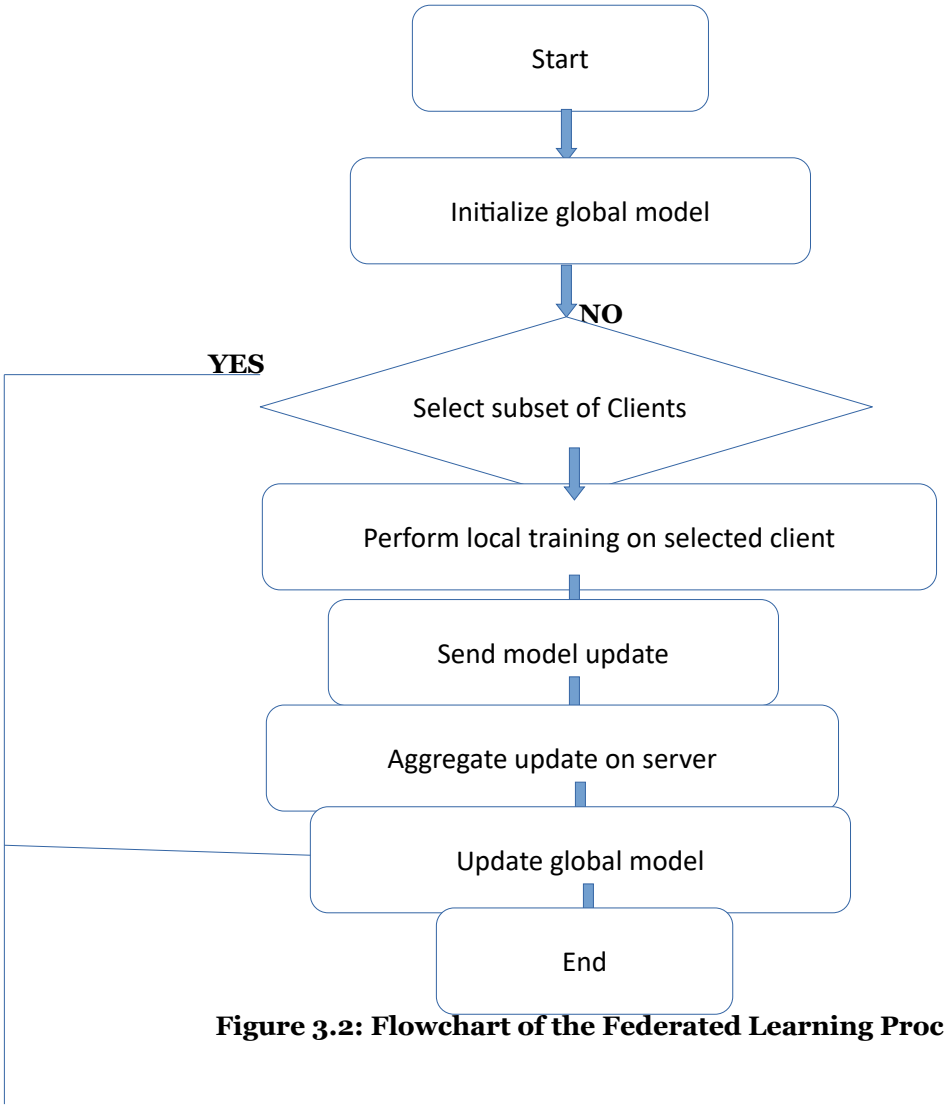

Figure 3.2: Flowchart of the Federated Learning Process

4. Implementation

Tools: For the implementation of Federated Learning (FL), TensorFlow Federated (TFF) is used, which provides the necessary framework to simulate federated training. The cloud simulation is carried out using CloudSim, a tool that allows for the representation and emulation of cloud environments, supporting in the analysis of system performance under various configurations.

Datasets: To evaluate the effectiveness of the proposed Federated Learning model, publicly available benchmark datasets such as MNIST and CIFAR-10 are utilized. These datasets are widely adopted in the machine learning community because they provide standardized benchmarks for image classification tasks. The MNIST dataset consists of grayscale images of handwritten digits from 0 to 9, while CIFAR-10 contains colored images categorized into ten distinct classes, including animals and vehicles. Together, these datasets offer diverse and representative data sources for training and evaluating FL models under different experimental settings [25][26].

4.1 Experimental Setup

- Metrics for Evaluation: To determine the efficiency of the FL model is evaluated through the use of these metrics:
- Model Accuracy: The primary metric for evaluating how well the model generalizes to unfamiliar data.
- Duration of Training: The overall time for training the model, which includes both local client training and global model aggregation.

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- Communication Overhead: The volume of information transferred between clients and the server, which is crucial for measuring the efficiency of the federated approach.
- Privacy Preservation Metrics: A qualitative measure of how effectively the model ensures the privacy of the data, often quantified through differential privacy or other privacy-preserving techniques.

4.2 Experimental Design

Client Configuration: The experimental setup involves simulating a varying number of clients to assess the scalability of Federated Learning. Configurations include 10, 20, and 50 clients To grasp the influence of client participation regarding the training and effectiveness of models. These different configurations will help simulate real-world scenarios in which the number of participating clients may vary.

Training Configuration: To study the effect of hyperparameters on model performance, different learning rates and local epochs are used. Learning rates are varied to observe how quickly the model converges, while local epochs (the number of times each client performs local training before sending updates) are altered to study their influence on both model precision and communication productivity [14]. This setup strives to offer an in-depth insight into how Federated Learning performs under different experimental conditions, ensuring that the results are applicable in various real-world scenarios.

5. Results and Discussion

This chapter presents and analyzes the results obtained from the simulated experiments conducted on the proposed Federated Learning (FL) framework. The goal is to evaluate the performance, privacy implications, and operational feasibility of FL when compared with traditional centralized machine learning techniques. By focusing on key performance indicators such as accuracy, training duration, communication overhead, and resource efficiency, this chapter provides a holistic understanding of the trade-offs and benefits of FL in real-world deployment scenarios. Furthermore, the discussion explores how the framework addresses core challenges such as data privacy, model convergence in non-IID settings, and scalability across heterogeneous edge-cloud infrastructures.

5.1 Performance Evaluation

To evaluate the practical implications of deploying Federated Learning (FL) in a cloud–edge environment, a comprehensive performance comparison was conducted between the proposed FL framework and a traditional centralized learning approach. This evaluation focuses on three critical performance dimensions, accuracy, training time, and communication overhead, which collectively determine the operational viability, efficiency, and data privacy impact of the two learning paradigms. In addition to these core metrics, the analysis also highlights the trade-offs associated with decentralized training, particularly the balance between slightly reduced accuracy and improved privacy preservation. Such an evaluation is essential to understanding not only the technical performance of FL but also its suitability for real-world deployment in data-sensitive industries.

Table 5.1: ***Performance Comparison of Centralized vs. Federated Learning***

Metric	Centralized Learning	Federated Learning
Accuracy (%)	95.0	93.7
Training Time (s)	120	140
Communication Overhead (MB)	60.0	12.0

5.1.1 Analysis and Discussion

- Accuracy: Centralized learning achieved a slightly higher model accuracy of 95.0%, compared to 93.7% attained by FL. This marginal decrease in FL accuracy can be attributed to the non-IID (non-independent and identically distributed) nature of data in federated settings. In centralized learning, all data is collected and aggregated in one place, allowing the model to learn from a diverse, balanced dataset. FL, however, trains locally on devices

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

where data may be class-imbalanced or context-specific. Despite this, FL's accuracy remains robust and sufficiently high for many practical applications, especially those where data privacy and decentralization are critical requirements.

- Training Time: The time taken to train the global model was 20 seconds longer in the FL setup. Centralized training completed in 120 seconds, whereas FL took 140 seconds. This increase stems from the overhead introduced by distributed computation, synchronization delays, and network-induced latency. Each FL training round requires multiple clients to train locally and communicate their model updates to a central aggregator. Despite the increase in time, the delay remains manageable and represents a fair trade-off for improved privacy and reduced data centralization.
- Communication Overhead: One of the most significant differences observed lies in communication overhead. FL drastically reduces data transfer requirements, consuming only 12 MB, compared to 60 MB in centralized learning—a reduction of 80%. In centralized systems, raw datasets from all participating clients must be transmitted to a central server. In contrast, FL shares only model updates (e.g., gradients or weight deltas), not raw data, making it ideal for bandwidth-constrained environments such as mobile networks, IoT deployments, and remote locations. This data minimization also mitigates potential attack surfaces and aligns FL with privacy regulations such as GDPR and HIPAA.

5.1.2 Visual Comparison

To further highlight the contrasts between the two learning paradigms, the comparative metrics are visualized in Figure 5.1. Figure 5.1: Bar chart comparing accuracy, training time, and communication overhead for centralized vs. federated learning.The chart clearly demonstrates that:

- Accuracy between the two paradigms is comparable, with FL showing only a minimal reduction.
- FL has a slightly higher training time due to distributed operations.
- FL dramatically outperforms centralized learning in communication efficiency, making it scalable and sustainable.

5.1.3 Summary and Implications

This comparative analysis underscores the fundamental trade-offs between traditional centralized training and the emerging FL paradigm:

- Centralized learning offers slightly higher accuracy and faster convergence but at the expense of data privacy and significantly higher communication costs.
- Federated Learning, while slightly slower and marginally less accurate, excels in privacy preservation, communication efficiency, and scalability.

These findings validate the use of FL in data-sensitive domains such as healthcare, finance, and personalized services, where data confidentiality, regulatory compliance, and edge-device compatibility are prioritized.

5.2 Performance Evaluation

The results of the experiments are presented in the following table, comparing Federated Learning with traditional centralized learning approaches.The results indicate that while Federated Learning may have slightly lower accuracy and longer training times, it significantly reduces communication overhead, making it a more efficient solution in terms of data privacy [8][13][14].

Table 5.2: Performance Comparison of Centralized vs. Federated Learning

Metric	Centralized Learning	Federated Learning
Accuracy (%)	95.0	93.7

Journal of Information Systems Engineering and Management
2025, 10(58s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Training Time (s)	120	140
Communication Overhead (MB)	60.0	12.0

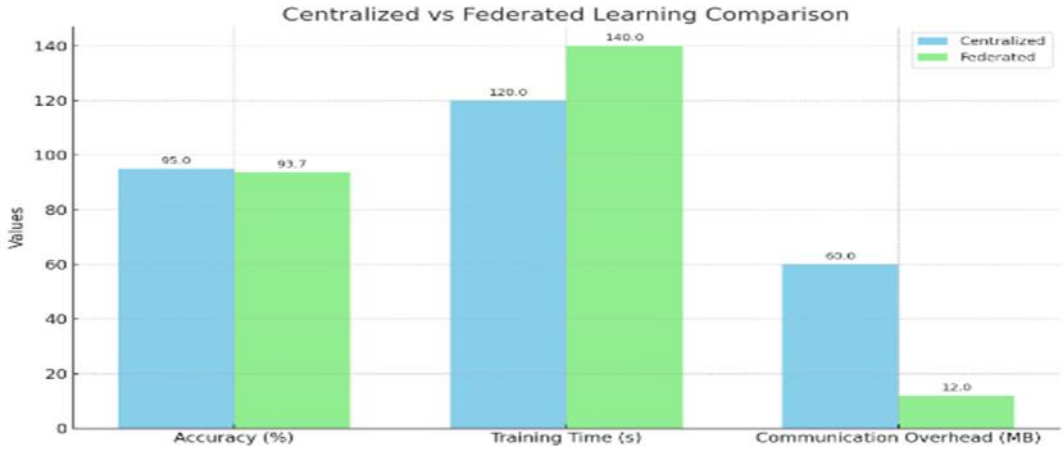

Figure 5.1: Centralized vs Federated Learning Comparison

5.2 Privacy Analysis

The integration of Federated Learning significantly enhances data protection. As information stays on the device, the likelihood of data breaches during transmission is minimized. This approach aligns with regulations for data protection like GDPR, which emphasize the significance of safeguarding data privacy [16].

5.3 Resource Utilization

The efficiency of resource usage in cloud environments when employing Federated Learning is analyzed. The reduced communication overhead in FL leads to lower bandwidth usage, making it a cost-effective solution for organizations. Additionally, the decentralized nature of FL allows for better utilization of local computational resources [8][13][14].

5.4 Comparative Analysis with Other Techniques

In comparison to traditional machine learning techniques, Federated Learning provides numerous benefits:

- Data Confidentiality: Sensitive information stays on personal devices, lowering the chance of exposure.
- Reduced Latency: Local training minimizes the need for data transfer, leading to faster model updates.
- Scalability: FL can efficiently scale with the number of clients, making it suitable for large distributed systems [18].

6. Conclusion

This research offers an in-depth investigation of the usage of Federated Learning (FL) within cloud computing environments. The proposed framework effectively demonstrates how FL can facilitate decentralized model training while maintaining user protection privacy by keeping sensitive information housed on local devices. Experimental evaluations show that although Federated Learning exhibits a minor trade-off in precision and extended training duration compared to centralized approaches, it drastically reduces communication burden and enhances adherence to data privacy laws like GDPR [8][13][14]. Moreover, the implementation of the Federated Averaging (FedAvg) method enables scalable, cognizant of privacy learning across distributed clients.

6.1 Implications for Future Research

The findings suggest several promising directions for future research. Further work may focus on refining FL algorithms through adaptive optimization strategies, including dynamic learning rates, personalized federated learning, and model compression techniques to reduce computational overhead. Moreover, the application of FL can be extended to complex and high-stakes domains such as smart cities, autonomous vehicles, remote diagnostics, and industrial IoT systems, where data sovereignty and responsiveness are essential [6][19][23]. Another important trajectory involves integrating blockchain with Federated Learning to enable secure, auditable model updates and decentralized trust management, thereby addressing challenges of transparency and accountability in large-scale deployments [22].

References

[1] M. Sardaraz and M. Tahir, "A Hybrid Algorithm for Scheduling Scientific Workflows in Cloud Computing," in IEEE Access, vol. 7, pp. 186137-186146, 2019.
[2] Abadi, M., Chu, A., & et al. (2016). Deep Learning with Differential Privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.
[3] Bonawitz, K., Hard, A., & et al. (2019). Towards Federated Learning at Scale: System Design. Proceedings of the 2nd SysML Conference.
[4] Chen, M., Zhang, Y., & et al. (2023). Federated Learning Meets Edge Computing: A Survey. IEEE Internet of Things Journal.
[5] Hard, A., Rao, K., & Mathews, R. (2018). Federated Learning for Healthcare: Model and Algorithm Design. Proceedings of the 2018 IEEE International Conference on Healthcare Informatics.
[6] Kairouz, P., McMahan, H. B., & et al. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning.
[7] Li, X., Zhang, Y., & et al. (2023). Federated Learning for Smart Homes: A Privacy-Preserving Framework. Journal of Network and Computer Applications.
[8] McMahan, H. B., Moore, E., & et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics.
[9] Yang, Q., Liu, Y., & et al. (2019). Federated Machine Learning: Concept and Applications. ACM Transactions on Intelligent Systems and Technology.
[10] Zhao, Y., Liu, Y., & et al. (2022). Federated Learning for Credit Scoring: A Privacy-Preserving Approach. IEEE Transactions on Knowledge and Data Engineering.
[11] Smith, V., et al. (2017). Federated Learning for Image Classification: A Case Study. Proceedings of the 2017 IEEE International Conference on Computer Vision.
[12] Kumar, R., Singhal, N., & Chhabra, A. (2025). Hybrid Optimization algorithm with the combination of PSO and genetic algorithm for task scheduling in cloud computing. E-Learning and Digital Media, 0(0). https://doi.org/10.1177/20427530251331082
[13] Huang, T., et al. (2020). Federated Learning with Non-IID Data. Proceedings of the 2020 International Conference on Machine Learning.
[14] Li, T., et al. (2020). Federated Learning: Challenges, Methods, and Future Directions. IEEE Transactions on Neural Networks and Learning Systems.
[15] Arpit Chhabra, Manav Bansal and Niraj Singhal, “Smart City-Shrewd Vehicle Versatility Utilizing IOT”, International Journal of Engineering Trends and Technology, Vol. 70, Issue 3, pp. 29-36, 2022.
[16] Wang, J., et al. (2021). A Survey on Federated Learning: From Theory to Applications. IEEE Transactions on Knowledge and Data Engineering.
[17] Wang, H., et al. (2022). Secure Federated Learning: A Survey. IEEE Transactions on Information Forensics and Security.
[18] Li, Y., et al. (2022). Federated Learning for Smart Healthcare: A Review. IEEE Access.
[19] Wang, S., et al. (2023). Federated Learning for Smart Cities: A Survey. IEEE Internet of Things Journal.
[20] Arpit Chhabra, Niraj Singhal and Syed Vilayat Ali Rizvi, “A Novel Algorithm of Safe-Route Traversal of Data for Designing the Secured Smart City Infrastructures”, International Journal of Engineering Trends and Technology, Vol. 71, Issue. 5, pp. 272-281, 2023.

[21] Wang, Y., et al. (2023). Federated Learning with Differential Privacy: A Survey. ACM Computing Surveys.
[22] Wang, Z., et al. (2023). Federated Learning for Cybersecurity: A Review. IEEE Transactions on Information Forensics and Security.
[23] Wang, Y., et al. (2023). Federated Learning for Autonomous Vehicles: A Survey. IEEE Transactions on Intelligent Transportation Systems.
[24] Zhang, Y., et al. (2023). Federated Learning in the Age of Big Data: A Survey. IEEE Transactions on Big Data.
[25] Kumar, R., Singhal, N., & Chhabra, A. (2025). Revolutionizing Business Management Strategies for Enhanced Output Through the Integration of Deep Learning and Cloud Computing. Journal of Information Systems Engineering and Management, 10(58s).
[26] LeCun, Y., Cortes, C., & Burges, C. J. C. (1998). *The MNIST Database of Handwritten Digits*. Available at: http://yann.lecun.com/exdb/mnist/
[27] Krizhevsky, A., & Hinton, G. (2009). *Learning Multiple Layers of Features from Tiny Images*. Technical Report, University of Toronto.