Acta Scientific COMPUTER SCIENCES

Volume 7 Issue 8 November 2025

Literature Review

Federated Learning - An Emerging Scenario of Future Data Science

V.J.K. Kishor Sonti[1]*, Sai Varun C[2], Sivasangari A[3] and A Ronald Doni[3]

[1]Professor, Department of ECE, Sathyabama Institute of Science and Technology, India

[2]Student, Department of CSE, Sathyabama Institute of Science and Technology, India

[3]Professor, Department of CSE, Sathyabama Institute of Science and Technology, India

***Corresponding Author:** V.J.K. Kishor Sonti, Professor, Department of ECE, Sathyabama Institute of Science and Technology, India.

DOI: 10.31080/ASCS.2025.07.0594

Received: September 23, 2025

Published: October 17, 2025

© All rights are reserved by **V.J.K. Kishor Sonti., *et al.***

Abstract

In today's world, privacy and security are the most important challenges to be addressed. The increasing number of interconnected devices has led to more data generation, creating many opportunities as well as vulnerabilities for Machine Learning (ML) applications. Decentralized Artificial Intelligent (AI) systems like Federated Learning (FL) have addressed all the issues related to data privacy and security. This is being achieved by introducing a secure distributed ML process, where there would be multiple nodes or local models and one global model or the server. This approach allows for multiple nodes to be trained individually on the training data and share a global model. This is trained by aggregating the weights obtained from all the nodes. A detailed view about the fundamentals of FL is presented in this work. The fundamentals of FL are explored with working principles, advantages and challenges faced with FL. Various case studies are provided about the usage of FL by various organizations to ensure data privacy and security. A detailed explanation about impact of FL and applications is presented. This also narrates the problems addressed, classifications of FL, challenges and its applications in various sectors. Additionally, the performance evaluation of FL is provided and is compared with the traditional AI systems by measuring accuracy, execution time, convergence time and Central Processing Unit (CPU) usage. This highlights the use of FL for ensuring privacy and security without affecting the performance of the AI model. The future scope section provides a comprehensive idea of further expansion of this concept of FL. This would be a comprehensive guide for academicians, researchers and enterprises as it offers an approach for ensuring data privacy and security without affecting the performance of the model. Various enterprises are struggling with privacy concerns, this chapter can guide them to employ this proposed approach to prevent any risk or data breach. This approach also ensure the safety of confidentiality data such as consumer data, employee data and all.

Keywords: Federated Learning; Case Studies; Confidentiality; Data; Privacy

Introduction

In recent years, there have been many data breaches and data leakages which increased the concerns of the public about data security [1-3]. Several organizations have suffered data breaches in recent years, the average loss to an organization due to a data breach in 2024 is $9.36 million and the financial loss is observed to be increasing at a range of 2.5-5% per annum. Additionally, these also affect the reputation of the organization along with legal issues and loss of customer trust.

Table 1 provides a comprehensive view of the average financial loss for organizations due to data breaches from 2014 to 2024, along with the annual increase percentage in costs. It is observed that the cost has increased from $3.52 million in 2014, to $4.88 million in 2024, highlighting an overall increase of approximately 38.64%. The highest annual increase is observed in 2021 (9.84%)

Table 1: Yearly Financial Loss and Annual Increase (%) in Global Data Breach Costs (2014-2024).

Year	Cost (in million USD)	Annual Increase (%)
2014	3.52	-
2015	3.79	7.67
2016	4.00	5.54
2017	3.62	-9.50
2018	3.86	6.63
2019	3.92	1.55
2020	3.86	1.53
2021	4.24	9.84
2022	4.35	2.59
2023	4.45	2.30
2024	4.88	9.66

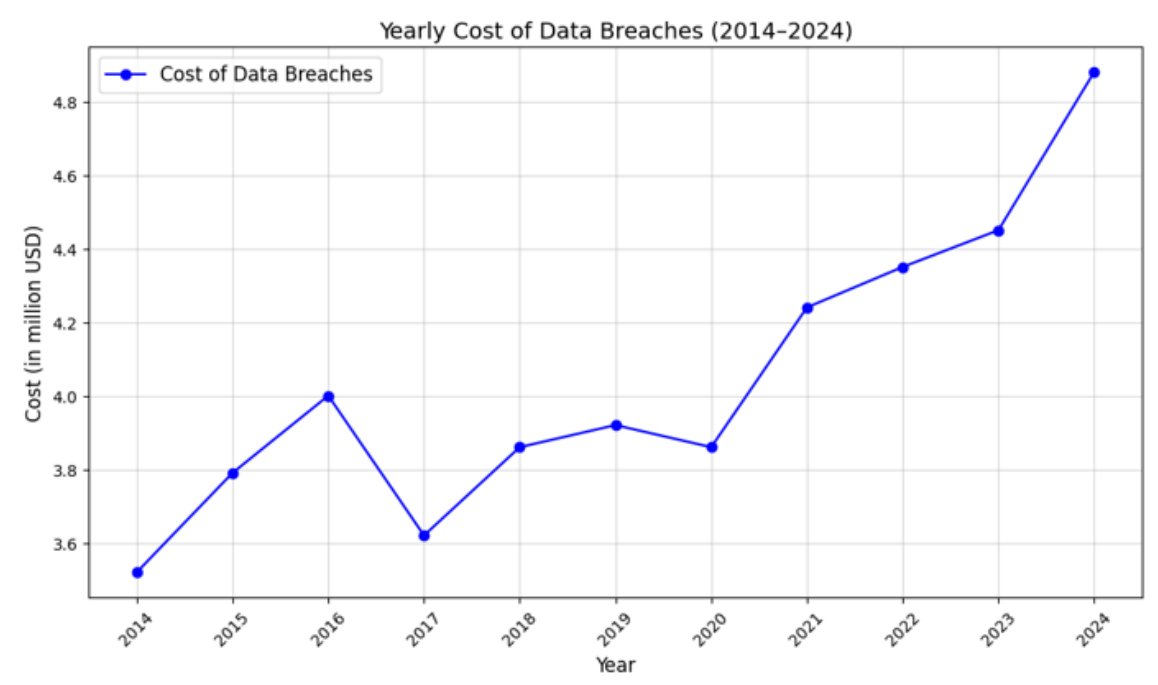

Figure 1: Yearly Cost of Data Breaches (2014-2024).

and 2020 (9.66%). Overall, the table has highlighted the gradual increase in financial loss due to data breaches in recent years.

Figure 1 illustrates the increasing trend in the global average cost of data breaches over a 10-year period from 2014-2024. The plot shows the increasing trend in costs, increasing from $3.52 million in 2014 to $ 4.88 million in 2024.

Literature

This section provides a comprehensive outlook about the existing literature leading to the evolution of FL. Data security is a prime concern for many organizations. Apart from organizations, individuals also face losses such as financial loss, identity theft and legal

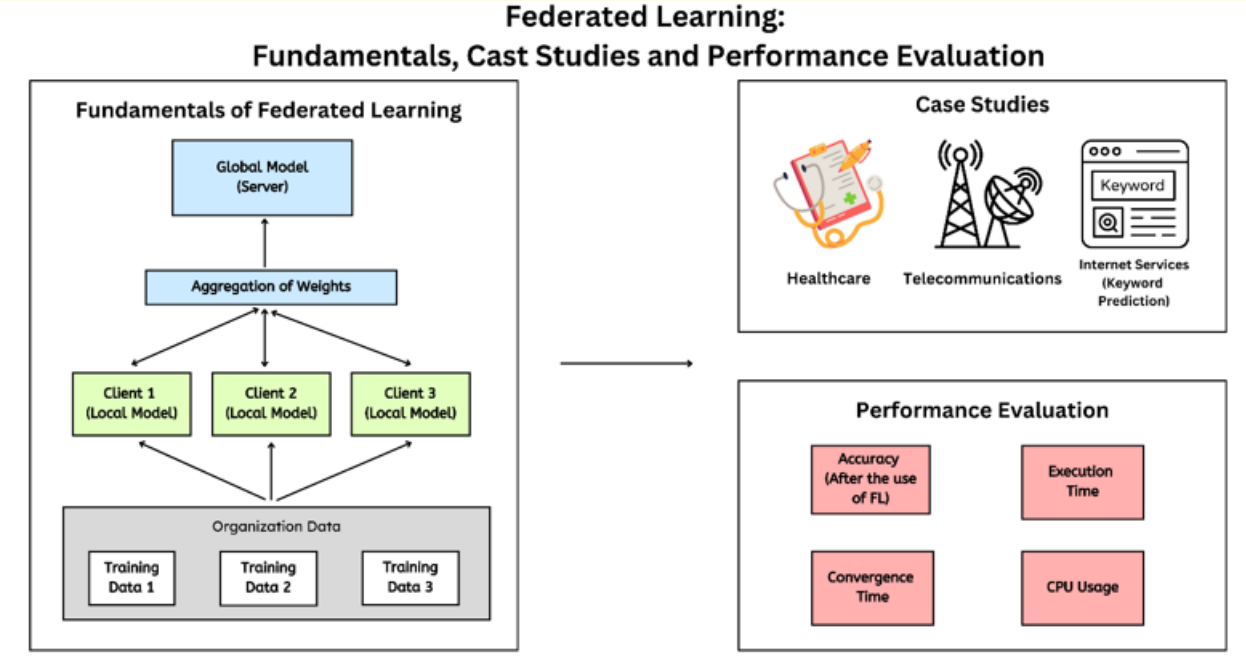

Figure 2: Federated Learning Overview.

issues. Many Individuals, groups and organizations are working together to improve data privacy and security [2,4]. Various regulations such as General Data Protection Regulation (GDPR) [5,6] in the European union, Health Insurance Portability and Accountability Act (HIPAA) for healthcare data in the USA and Digital Personal Data Protection (DPDP) are introduced to ensure data privacy and security but are not sufficient to address the issue as these usually outline only the general principles and rules [7]. The lack of focus on technical details and emerging threats leaves the room for attackers to exploit vulnerabilities beyond the scope of these regulations. The rise of interconnected devices has led to more data generation, which has created new opportunities for ML applications as AI systems mainly rely on the data collected and processed [8]. This has resulted in more data breaches and leakages.

Centralized AI systems are the common approaches that cause data privacy and security risks, as during the model training, the data is exposed to the world. These challenges paved the way for

the rise in decentralized AI systems, where there is no centralized model training. The decentralized AI systems enable computation directly on the local devices instead of the centralized systems which depend on training large amounts of data in a centralized server. Here, the server aggregates the weights obtained from the local model to update the global model. The most used decentralized AI system is FL.

Current scenario of FL

Nowadays, there is a growing demand for FL systems, as more companies across various sectors are adopting them to ensure data privacy and security. The global FL market is expected to grow at a Compound Annual Growth Rate (CAGR) of 10.2% from 2022 to 2032, due to the increasing demand for decentralized AI systems. The market is expected to reach 311.4 million USD in 2032. Leading companies like Google LLC, IBM Corporation, Nvidia Corporation, Intel, FedML, Accuratio, Edge Delta, Secure AI labs and many others are contributing to the growth of FL.

Figure 3 shows the expected growth of the global FL market from 2017 to 2032 with the market size values in USD millions. The x-axis represents the years, and the y-axis represents the market size in USD. The chart shows a gradual increase in market size over the years with significant growth starting from 2017.

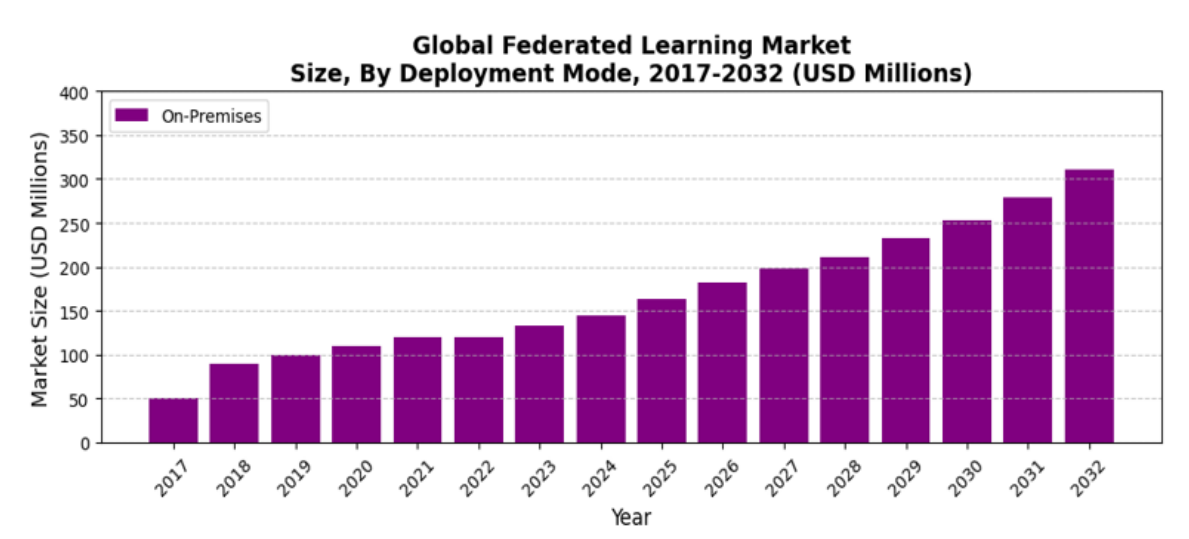

Figure 3: Global FL Market Size (Year vs Market Size).

FL is a distributed ML approach and an architecture in which the models are trained directly on the local devices and the global model is updated by aggregating the weights collected from all the devices. It basically involves sending the weights from the local model to the global model and then the server aggregates all the received weights to update the global model. This process is repeated indefinitely until the model achieves a predefined performance metric such as target accuracy or minimal loss value. Here, it is observed that FL enables building an ML model without sharing the client's data, ensuring data privacy and security [9-11]. For example, let's consider that a user wants to train an ML model on their smartphone in a centralized approach, then the main challenge that needs to be addressed would be the lack of data available for a central server to train the model and achieve higher accuracy. FL outperforms in such circumstances as it allows the transfer of weights from the smartphone to the global model for aggregat-

ing all the weights and updating the global model, benefiting all the clients or users. There are also various types of FL such as transfer learning, horizontal and vertical. The horizontal FL involves the same features but different samples for various clients. Similarly, vertical FL involves different features with the same sample space. Transfer FL is utilized when there is insufficient data to train a ML model.

The primary advantages of this type of AI system is that less latency is achieved, and it also maintains data privacy and security [8,12-15]. Additionally, FL also allows the user from different geographic locations to collaboratively train the ML model, and it doesn't compromise on the user privacy or violate any law or regulation. This highlights FL as superior in comparison with conventional AI systems. There are various sectors such as healthcare [16], internet services [17], telecommunications [18] or social networks [19] which have employed FL for building ML models with ensuring data privacy and security. The most common use cases of FL are keyword prediction in smartphone keyboards and speech recognition in AI assistants. However, FL has few vulnerabilities involve [20,21], which require privacy-preserving techniques such as differential privacy, homomorphic encryption [22] and secure aggregation to be implemented in the FL environment. Many organizations such as Google have employed secure convergence and differential privacy to ensure the privacy of data and security in the FL system [23].

Conclusion and Scope for Research

The primary focuses of this work are as follows:

- The aim is to explore the fundamentals of FL with key concepts such as working principles, advantages and challenges faced with FL. It also provides detailed overview of FL, classifications of FL and its applications in various sectors.
- Several case studies are provided about the usage of FL by various organizations to ensure data privacy and security. Furthermore, discussions are provided about the existing systems that employed FL for classification.
- Additionally, the performance of FL is provided and compared with the traditional AI systems by measuring accuracy, execution time, convergence time and CPU usage. This highlights the working of FL without affecting the performance of the AI model.
- With the enormous literature available on Machine learning and Deep learning (for ex: significant contributions of Yarici, Yurdem and others), FL has its scope to play in future research. The role of FL is more evident in the areas of providing security to data and is promising for nextgen systems protection. More usage of FL also considerably reduces the cost in maintenance of security to data.

Bibliography

1. Gong M., *et al*. "Privacy-enhanced multi-party deep learning". *Neural Networks* 121 (2020): 484-496.

2. Xie Y., *et al*. "Secure collaborative few-shot learning". *Knowledge-Based Systems* 203 (2020): 106157.

3. Bonawitz K. "Towards federated learning at scale: System design". arXiv preprint arXiv (2019): 1902.01046.

4. Zhang C., *et al*. "A privacy-preserving multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition". *Frontiers in Neurorobotics* 13 (2020): 112.

5. Albrecht JP. "How the GDPR will change the world". *European Data Protection Law Review: EDPL* 2 (2016): 287.

6. Regulation P. "Regulation (EU) 2016/679 of the European Parliament and of the Council". *Regulation* (eu) (2016): 679.

7. Boban M. "Digital single market and EU data protection reform with regard to the processing of personal data as the challenge of the modern world". *Economic and Social Development: Book of Proceedings* (2016): 191.

8. Yazici İ., *et al*. "A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems". *Engineering Science and Technology, an International Journal* 44 (2023): 101455.

9. Konečný J., *et al*. "Federated optimization: Distributed machine learning for on-device intelligence". arXiv preprint arXiv (2016): 1610.02527.

10. Kairouz P., *et al*. "Advances and open problems in federated learning". *Foundations and Trends® in Machine Learning* 14.1–2 (2021): 1-210.

11. Shokri R and Shmatikov V. "Privacy-preserving deep learning". In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security (2015): 1310-1321.

12. Yurdem B., *et al*. "Federated learning: Overview, strategies, applications, tools and future directions". *Heliyon* (2024).

13. Gong M., *et al*. "A survey on differentially private machine learning". *IEEE Computational Intelligence Magazine* 15.2 (2020): 49-64.

14. Zhang C., *et al*. "A survey on federated learning". *Knowledge-Based Systems* 216 (2021): 106775.

15. Li T., *et al*. "Federated learning: Challenges, methods, and future directions". *IEEE Signal Processing Magazine* 37.3 (2020): 50-60.

16. Xu J., *et al*. "Federated learning for healthcare informatics". *Journal of Healthcare Informatics Research* 5 (2021): 1-19.

17. McMahan B., *et al*. "Communication-efficient learning of deep networks from decentralized data". In Artificial intelligence and statistics (2017): 1273-1282.

18. Flanagan A., *et al*. "Federated multi-view matrix factorization for personalized recommendations". In Machine learning and knowledge discovery in databases: European conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part II (2021): 324-347.

19. Li T., *et al*. "Fair resource allocation in federated learning". arXiv preprint arXiv (2019): 1905.10497.

20. Blanco-Justicia A., *et al*. "Achieving security and privacy in federated learning systems: Survey, research challenges and future directions". *Engineering Applications of Artificial Intelligence* 106 (2021): 104468.

21. Kairouz P., *et al*. "Advances and open problems in federated learning". *Foundations and Trends® In Machine Learning* 14.1–2 (2021): 1-210.

22. Liu Y., *et al*. "A secure federated transfer learning framework". *IEEE Intelligent Systems* 35.4 (2020): 70-82.

23. Bonawitz K., *et al*. "Practical secure aggregation for privacy-preserving machine learning". In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (2017): 1175-1191.