Federated learning: Types, Algorithms, Applications, and Future Scope

Reshma Jadhav[1], Sarita Patil[2]
[1]*Reshma Jadhav, DYPCOE, Pune*
[2]*Sarita Patil, SICA, Pune*

***Abstract*—This paper provides a detailed review and analysis of privacy-preserving techniques in Federated Learning (FL). In today's world everyone needs data privacy and security. As observations Traditional ML algorithms, it is difficult to preserve sensitive data during their training phase. FL ensures that user data remains confidential during the training and testing phase. The paper discusses the advancements in secure aggregation, homomorphism encryption, and differential privacy within the context of FL. Furthermore, we address the challenges and trade-offs associated with these privacy-preserving mechanisms and propose potential avenues for future research in making FL more privacy-friendly. The goal is to contribute to the ongoing dialogue on striking the right balance between model performance and user privacy in the era of decentralized machine learning.**

***Index Terms*— FL, HE, Privacy, decentralized ML, Central Server, FedAVG, FedSGD**

I. INTRODUCTION

What is federated learning?

Federated learning is a decentralized method of training machine learning models (often called collaborative learning). Communication of data from client devices to global servers is not necessary. Rather, the model is trained locally using the unprocessed information on edge devices, thus enhancing data privacy. The local improvements are aggregated to build the final model in a shared way. [3].Here are some reasons why federated learning matters.

1. Privacy: Federated learning allows for local training on the edge device, avoiding potential data breaches, as opposed to traditional systems that send data to a central server for training.
1. Data Security: To ensure data security, only those encrypted modifications to the model are communicated to the central server. In addition, only aggregated results may be decrypted using safe aggregation approaches like safe Aggregation Principle.
2. Access to heterogeneous data: Access to data diffused across several devices, locations, and organizations is ensured via federated learning. It allows for secure and private training building models on sensitive data, like financial or medical data. Additionally, models may be made more generalizable with increased data diversity.

How does federated learning work?

Federated learning doesn't need data to reside on a central server to train a machine learning model. Federated learning trains centralized models using decentralized methods. Because federated learning is iterative, training may be done multiple times, allowing for ongoing learning and knowledge exchange.

- Selecting a model either pre-trained or not at all, is the first stage. Next, the initial model is sent to local devices or local servers.
- Local datasets are used to train machine learning models locally.
- The cloud has access to the local models.
- A globally shared model is constructed.
- Aggregate values are used by global models to determine optimal performance.
- Subsequently, local data centers get attributes from global models in order to incorporate global models into local models.

A global sharing approach facilitates cooperative learning across several devices. It updates the model using the data on your devices and only transmits the model's acquired information (such as parameters and outcomes) to the cloud not the data itself. It means that by keeping data local, it secures individual data.

For instance, while protecting the privacy of your text message, the keyboard model uses the sharing model to anticipate the word that will come next for you. By storing your data locally, this decentralized machine learning technology lowers the need of hardware infrastructure. [1]

The following steps make up the basic federated learning process, as shown in Figure 1:

1. The global model is downloaded by the participant from the server.
2. To obtain a local model, participant n trains local data.
3. Local model changes are uploaded by participants to the central server.
4. The global model is obtained by the server using a weighted aggregation process once it has received the data from all participants.

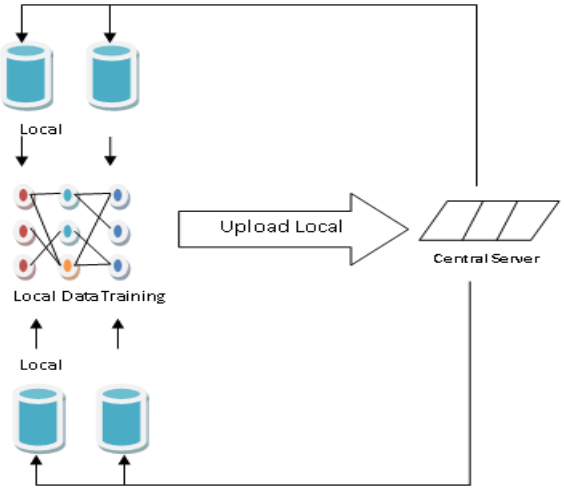


Fig1. Federated Learning Process

The federated learning technology has the following features:

1. Only model update data necessitates communication with a centralized server; all other data used in federated learning are always kept locally on clients and are never communicated in plain text.
2. Everybody who provides data for the model's alpha training will be able to access the final model as part of federated learning.
3. In the end, the accuracy of a federated learning model is equivalent to that of a centrally centralized machine learning model.
4. As the quality of training data utilized by the federated learning users rises, so does the global model's accuracy.

II. LIFE CYCLE OF FEDERATED LEARNING MODEL

Federated Learning is a process of Model deployment for particular application.

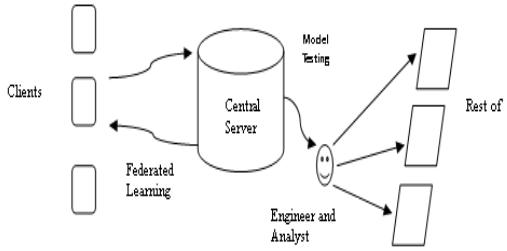
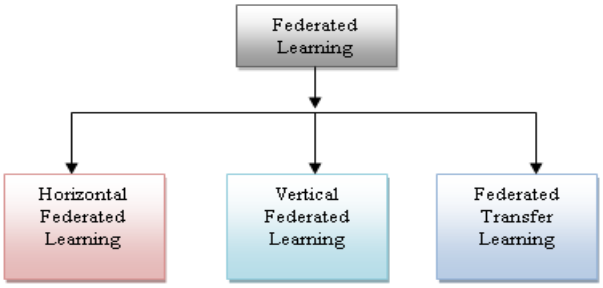


Fig.2 Federated Learning Lifecycle model

1. Problem Identification: The Engineer will identify the problem to be solved with FL
2. Client Selection: select local devices that could take part in training and testing phase.
3. Model Training: Central server sent a model to local server and local server train the model on their own dataset and sends back the updated model to central server.
4. Model Testing: Central server update its model and test it once it will reach to model convergence this model deploys for rest of the world [9].

III. TYPES OF FEDERATED LEARNING

Federated learning has demonstrated significant promise in contexts where privacy is a concern, like the banking industry, manufacturing, and other domains related to data perception. Federated learning may be categorized according to the distribution of data are as:



1. Horizontal Federated Learning:

Data set integration is the foundation of horizontal federated learning. The high degree of similarity between user and participant data suggests that this is an area of interest. The subset of data where both participants have the same properties, despite the

consumers being different, is the data that may be utilized for collaborative model training (Figure 3). There is a wider variety of potential scenarios for horizontal federated learning applications in terms of data.

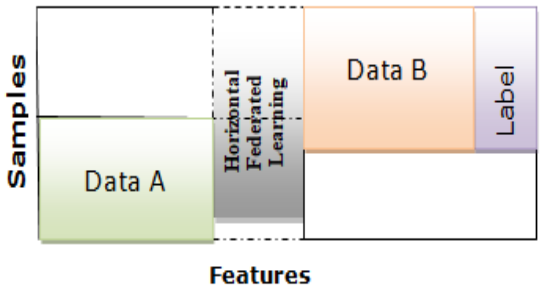


Fig. 3. Horizontal Federated Learning [1]

2. Vertical Federated Learning:

Through vertical federated learning, parties possessing disparate qualities pertaining to a same user base can collaborate to train machine learning models without disclosing their raw data or model parameters. The subset of data that differs in the context of data characteristics for the same users on both sides can be used for joint modeling training when there is greater overlap between users and less overlap between data features (Figure 4).

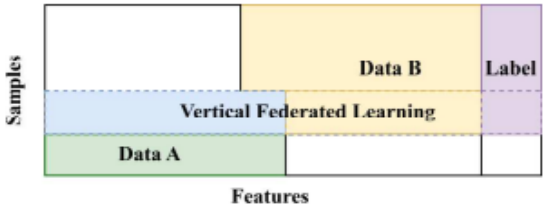


Fig.4. Vertical Federated Learning [1]

3. Federated Transfer Learning

In general, longitudinal data does not fill the same feature space or sample area as it does in the situations of vertical and horizontal federated learning. Therefore, the lack of data labels and their poor quality are the main problems here. When participant samples and attributes are not very similar, federated transfer learning can be helpful (Figure 5). The idea behind federated transfer learning is that each member has certain qualities.

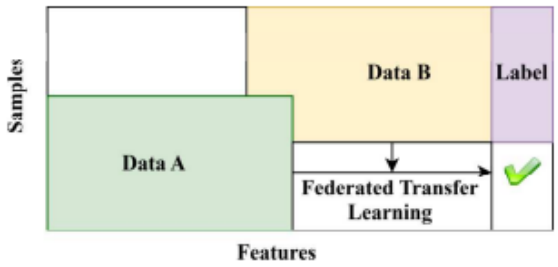


Fig.5. Federated Transfer Learning [1]

IV. Different Federated Learning Algorithms

1. Federated Averaging (FedAVG) :

This mechanism is used to train the data which is spreader across various different Servers or Devices; also it ensues not sharing of row data for privacy and security preservation

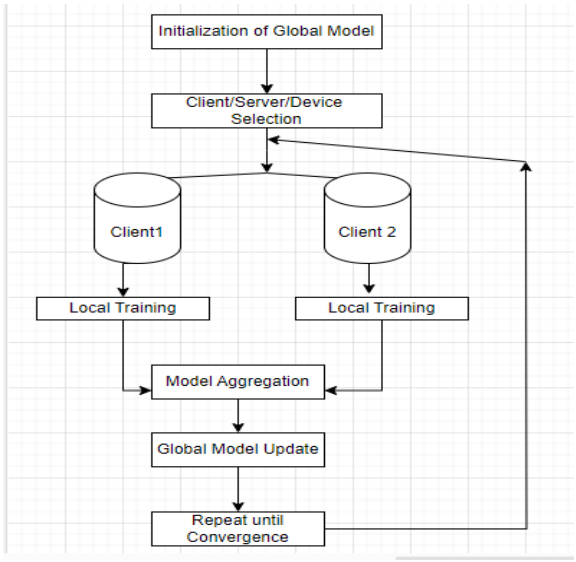


Fig.6 FedAvg Working Mechanism

1. Initialization: Central Server initialized a global model
2. Client Selection: Subset of client has been chosen to take part in training round
3. Local Model training: the global model sent to local devices and each local device train the module using its local dataset.
4. Model Aggregation: The trained models from each local device are sent back to the central server.
5. Repeat from step 2 for multiple training until convergence.

2. Federated Stochastic Gradient Descent (FedSGD)

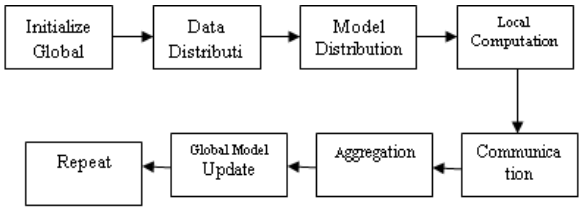


Fig.7 FedSGD Working Mechanism

1. Initialization: Central Server initialized a global model
2. Data distribution on local devices: Distribute training data on local devices, each device holds sample of its own dataset
3. Model Distribution: Send current global model to all local devices
4. Local Computation: Compute Gradient on each local device using stochastic Gradient Descent or any other similar optimization algorithm.
5 Communication: Local Devices sent the Computed Gradient Back to the central server
6. Aggregation: Aggregate the gradients received from local devices to central server
7. Repeat: Repeat the process for certain number of times until convergence.

3. Homomorphic Encryption Based FL:

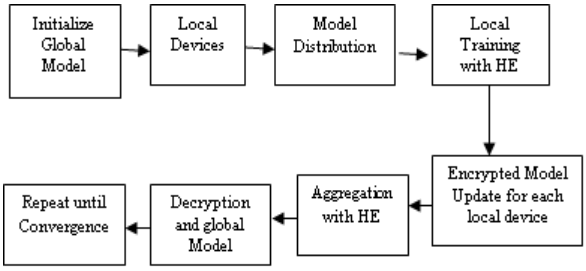


Fig.8 HE Working Mechanism

1. Initialize: Initialize global model on central server
2. Model Distribution: The initialized global model sent to local devices.
3. Local Training with HE:
A. each device performs local model training on its own dataset
B.homomorphic encryption applied on each local device data allowing computations on encrypted data without decrypting.
4. Aggregation with encryption: Central server receives aggregated encrypted data from local devices
5. Decryption and global model update: Central server decrypts the aggregated data and updates the central global model.
6. Repeat the process for certain number of times until convergence.[7]

V. APPLICATION OF FEDERATED LEARNING

Federated Learning (FL) is a machine learning technique that allows the training of models without data exchange among decentralized devices or servers that store local data samples. There are uses for this privacy-preserving method in many different fields. The following are some possible uses for Federated Learning:

1. Smartphone:

Statistical models use user behavior collected from a large number of mobile phones to power applications such as voice recognition, facial identification, and next-word prediction. To maintain their privacy, save battery life, or limit internet usage on their phone, users can choose not to share their data. With federated learning, precise Smartphone predictions may be made without affecting user experience or disclosing private information.

2. Organization:

In federated learning, entire institutions or organizations may be seen as "devices". For instance, hospitals have enormous amounts of patient data that might be utilized in apps for predictive healthcare. Hospitals, on the other hand, have strict privacy regulations in place and could be constrained by ethical, administrative, or regulatory requirements that need local data storage. For these applications, federated learning works well because it lowers network demand and enables private learning across several devices and organizations.

3. IoT:

Sensors are utilized in current IoT networks, such wearable technology, driverless vehicles, and smart homes, to collect and process data in real time. For example, a fleet of autonomous vehicles could need the most recent traffic, construction, or pedestrian behavior models in order to function properly. However, because of privacy concerns and each device's limited connectivity, creating aggregate models in these situations can be challenging.

Federated learning techniques allow us to train models that efficiently adapt to changes in these systems without compromising user privacy.

4. Healthcare:
Healthcare is one of the sectors that can gain the most from federated learning since sensitive health information cannot be shared readily owing to Health Insurance Portability and Accountability Act and other constraints. With this method, a significant quantity of data from different healthcare databases and gadgets may be utilized to build AI models that adhere to regulations.

5. Advertising:
As you are aware, personalization is heavily reliant on the information about each individual user. However, websites like social networking, ecommerce platforms, and other places spring to mind when more people worry about how much information they would prefer not to share with others. Advertising, which depends on customer personal data, may employ federated learning to stay afloat and allay people's fears.

6. Autonomous Vehicles:
Federated learning is used in the development of self-driving cars since it can make predictions in real time. The data may contain real-time updates on the state of the roads and traffic, facilitating quicker decision-making and ongoing learning. This might lead to a safer and more pleasurable experience when driving a self-driving car. One potential use of federated machine learning is in the automotive industry. But as of right now, research is the only activity being conducted in this area. Federated learning may reduce training time for self-driving automobile wheel steering angle prediction, according to one study.

7. Federated learning in the field of financial fraud:
The digital era has given rise to several international financial crimes. Financial crimes frequently fall under the following subcategories: money laundering, fraudulent lending, and financial crimes. Credit card fraud causes big costs for banks and customers.

8. Federated learning in the field of insurance:
In the process of creating a data service platform for the insurance industry, financial, medical, and other data from several sources must be integrated. If an insurance company wants to improve both its business development and risk management capabilities, multi-party data must be taken into account. In the insurance industry, efficient data utilization that respects individual privacy is a major challenge as well.[3]

VI. CONCLUSION

In this paper we have discussed FL as decentralized method for training ML models with ensuring of data privacy, data security and heterogeneous data access. We explore fundamental concept of FL, its life cycle, and various types FL such as horizontal vertical and federated transfer learning. Furthermore, we have discussed various FL algorithms including FedAvg, FedSGD. also this paper focuses on various FL applications such as Smartphone, healthcare etc.
In conclusion Federated Learning holds immense promise as data privacy and collaborative approach to ML. In future this technique is widely used across various sectors where we need secure data processing with removing challenges such as training ML model on various client devices such as mobile phones.

REFERENCE

[1] Sonam Tyagi, Ishwari Singh, and Richa Pandey, “Federated Learning Applications,Security Hazards and Defence Measure ”, Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551,023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT) | 978-1-6654-7491-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/DICCT56244.2023.10110075.

[2] Jie Wen, Zhixia Zhang, Yang Lan, Zhihua Cui,Jianghui Cai,Wensheng Zhang, ” A survey on federated learning: challenges and applications “ , International Journal of Machine Learning and Cybernetics (2023) 14:513–535.

[3] Subrato Bharati, M. Rubaiyat Hossain Mondal, Prajoy Podder, V.B. Surya Prasath, “ Federated learning: Applications, challenges and future directions “ , International Journal of Hybrid Intelligent Systems 18 (2022) 19–35 19 DOI 10.3233/HIS-220006.

[4] Abdulkadir Korkmaz, Ahmad Alhonainy, Praveen

Rao," An Evaluation of Federated Learning Techniquesfor Secure and Privacy-Preserving Machine Learning on Medical Datasets", 2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) | 978-1-6654-7729-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/AIPR57179.2022.10092212.

[5] Zilinghan Li, Shilan He, Pranshu Chaturvedi, Trung-Hieu Hoang, "APPFLx: Providing Privacy-Preserving Cross-Silo Federated Learning as a Service "2023 IEEE 19th international conference on e-Science 8979-8-350-3-2223-1/23/$31.00 @2023IEEE.

[6] Harinath Bodagala, Dr. Priyanka H, "Security for IoT using Federated Learning", 2022 International Conference on Recent Trends in Microelectronics.

[7] Yixuan Zhao, "Comparison of Federated Learning Algorithms for Image Classification", 2023 2nd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI) | 979-8-3503-3976-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDACAI59742.2023.00122.

[8] Tao Sun, Dongsheng Li, Bao Wang, " Decentralized Federated Averaging", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 4, APRIL 2023.

[9] Peter Kairouz, H. Brendan McMahan, Brendan Avent," Advances and Open Problems in Federated Learning", arXiv:1912.04977v3 [cs.LG] 9 Mar 2021.