Federated Learning in the Healthcare Sector: Opportunities and Difficulties

Indraneel Mukhopadhyay, Abhinandan Ghosh
Amity Institute of Information Technology,
Amity University Kolkata
imukhopadhyay@gmail.com

ABSTRACT

Under the supervision of a central server, many devices cooperatively develop a machine learning model without disclosing their personal data, a concept known as federated learning (FL), which was initially presented by Google. This presents several potentials in vital industries like healthcare, banking, etc., where disclosing personal user data to other entities or gadgets is dangerous. Although FL seems like a promising Machine Learning (ML) method for protecting the privacy of local data, it is susceptible to assaults much like other ML models. This paper addresses federated learning's prospects and problems considering the field's increasing attention.

1. INTRODUCTION

In the recent four to five years, artificial intelligence (AI) and machine learning (ML) have gained popularity. The use of Machine Learning technologies in several sectors, including banking, healthcare, transportation, customer services, e-commerce, and smart home applications, has been expedited by the availability of Big-data and powerful processing units. Because machine learning techniques are being used widely, it is crucial to protect the techniques' security and privacy. The majority of machine learning applications combine data from different devices or organizations into a single server or cloud platform in order to train the model. This is a serious drawback, particularly when there are security risks in the training data set because to the sensitive information it includes.

Google unveiled the concept of Federated Learning, which was initially used in the Google Keyboard to facilitate collaborative learning from several Android phones [1]. Because FL may be implemented on any edge device, it has the potential to completely transform a number of important industries, including banking, healthcare, transportation, and smart homes. The most well-known instance is when scientists and doctors from several regions of the world collaborated to create an AI pandemic engine for COVID-19 diagnosis using chest scans [2]. Transportation networks provide an intriguing application for teaching autonomous driving and city route planning to automobiles. In a similar vein, federated learning frameworks enable edge devices in various houses to cooperatively learn on context-aware rules for smart-home applications [3].

Federated learning presents a number of obstacles despite its wide range of applications. The issues may be generally divided into two categories: security challenges and training-related challenges. The obstacles associated with training include heterogeneity of the devices involved in the learning process, heterogeneity of the data utilized for training, and communication overhead over several training iterations. On the other hand, security concerns encompass risks to privacy and security arising from adversaries, which might range from hostile users with black-box access to the model to harmful clients present in the local device.

Even while private data in FL stays on the device, it might still be feasible for an enemy or an

inquisitive bystander to discover that a data point used for training in the local models exists. To prevent this attack, the information must be kept differentially private using some sort of cryptographic mechanism [4]. On the other hand, malevolent clients present in the learning environment have the ability to primarily trigger security assaults, which can be either non-targeted or targeted. The goal of targeted attacks is to influence the labels on certain tasks. On the other hand, in non-targeted assaults, the adversary's only goal is to undermine the global model's correctness. Defense measures must identify hostile devices, stop them from learning more, or neutralize the impact they have on the global model [5].

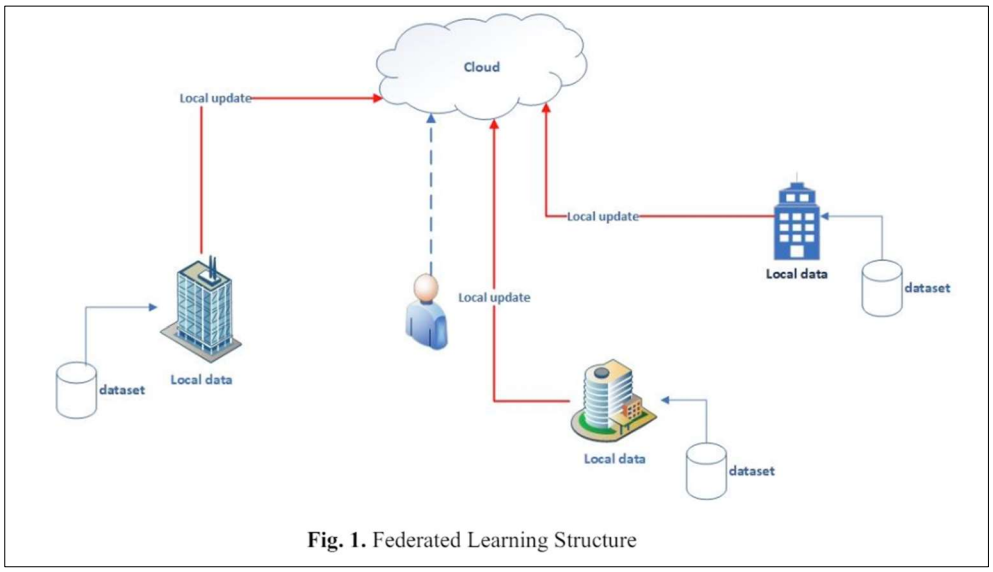

Fig. 1. Federated Learning Structure

2. TYPES OF FEDERATED LEARNING

This section presents several Federated Learning framework types.:

1. Vertical Federated Learning - When each device has a separate dataset with sample instances but different attributes, vertical federated learning is employed. For example, Vertical FL may be used to develop a shared machine learning model when two businesses have distinct feature sets but the same collection of data about the same individuals.
2. Horizontal Federated Learning - When each device has a dataset with the same feature space but distinct sample instances, horizontal federated learning is employed. This kind of learning is used in the first FL-Google keyboard use case, wherein the participating mobile phones have separate training data but the same characteristics.
3. Federated Transfer Learning - Federated Transfer learning is comparable to regular machine learning in that it involves adding a new feature to a model that has already been trained. The greatest example would be to expand vertical federated learning if we wanted to use machine learning to a larger number of sample instances that aren't available in every cooperating company.
4. Cross-Silo Federated Learning - Coordinated Cross-Sector Learning is implemented when there are fewer participating devices and they are available for each round. For the training set, FL formats in both vertical and horizontal orientations are allowed. Cross-silo is mostly applied in organizational contexts. Cross-silo FL is used to construct the model in publications [6].

3. FEDERATED LEARNING IN HEALTHCARE

Federated learning is an emerging model in distributed learning which addressed these challenges by allowing collaborative training of ML models across multiple decentralized data sources without the requirement to share the raw data. In the field of healthcare, federated learning can drastically enhance the capabilities of Computational Intelligence and Neuroscience by providing a framework for integrating insights from diverse medical datasets. This collaborative approach is further expanded by technologies such as Model Deployment Management as a Service (MDMaaS), which streamlines the deployment and management of AI models in clinical settings. Moreover, advancements in algorithms like communication-efficient parallel Stochastic Gradient Descent (cpSGD) are crucial for optimizing the training process in federated environments, ensuring that models are both accurate and efficient.

This process makes sure that the patients' privacy is maintained. To achieve this it uses differential privacy, which adds noise to the data to prevent extraction of information. FL supports the computational power of multiple institutions, thereby supporting the creation of robust AI model while sticking to strict data protection regulations. The application of AI in the healthcare has changed the industry. Now we have new paths of diagnosis, treatment and patient care. machine learning techniques like Convolutional neural network have demonstrated greater efficacy in image-based diagnosis such as radiology and pathology. but there are certain hinderances such as privacy concerns, data security issues, and the logistical challenges of aggregating vast amounts of sensitive patient data from diverse sources.

This paper investigates the application of federated learning in healthcare emphasizing its potential to overcome current challenges to AI adoption. By leveraging federated learning, healthcare institutions can collaboratively build and refine AI models. This enhances diagnostic accuracy and improves treatment outcomes. It also maintains patient privacy and data security at the same time. The discussion includes a detailed examination of methodologies used and the advantages and challenges of federated learning in healthcare. Additionally, there is future scope of this transformative approach.

4. LITERATURE SURVEY OF HEALTHCARE SOLUTIONS

Federated Learning (FL) in healthcare can be incorporated in numerous ways, particularly for enhancing predictive modeling and disease classification while ensuring patient privacy. FL facilitates collaborative model training across various institutions without the need to share sensitive data, effectively addressing the privacy issues inherent in traditional machine learning (ML) methods. This decentralized approach proves particularly effective in several critical tasks, including medical image analysis, smart healthcare applications, and disease outbreak prediction.

Advanced techniques such as Differential Privacy and NbAFL further bolster FL by introducing noise to data updates, thus mitigating the risks of information leakage and enhancing overall data security. FL's utility extends to specific healthcare scenarios, such as predicting COVID-19 patient mortality and utilizing Electronic Health Records (EHRs) for forecasting cardiac events. In these contexts, FL enables the aggregation of valuable insights from diverse datasets without compromising the confidentiality of patient information. However, despite its advantages, FL faces several challenges. Data heterogeneity, where data varies significantly across different sources, poses a significant hurdle. Communication efficiency is another concern, particularly when devices have varying connectivity levels. Additionally, integrating various data modalities—such as combining textual EHR data with imaging data—requires sophisticated techniques to

ensure seamless and accurate analysis.

Despite these challenges, the promise of FL in healthcare remains substantial. Its ability to provide secure, collaborative data analysis can lead to significant improvements in healthcare outcomes. By enabling institutions to leverage collective data insights while maintaining strict privacy standards, FL paves the way for more effective and secure healthcare solutions. Ongoing research and development efforts continue to refine FL techniques, aiming to overcome existing technical and security obstacles and fully realize FL's potential in revolutionizing healthcare data analysis.

Reference	Methodology	Advantage	Disadvantage	Future scope
A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics, by K. Dasaradharami Reddy, Thippa Reddy Gadekallu [7]	This research leverages federated learning (FL) to train machine learning models for healthcare applications, addressing data privacy. FL allows training on local data from multiple sources without sharing raw data, enhancing security. The study explores machine learning for tasks like image analysis and diagnosis, while also discussing collaborative AI facilitated by FL and blockchain integration for improved data quality, model accuracy, and security. Homomorphic encryption further protects sensitive data throughout the process.	Decentralized Training: FL enables decentralized training models, facilitating collaboration between multiple parties without centralizing data, leading to improved data security and reduced risks of data breaches.	Data heterogeneity in healthcare settings poses challenges for Federated Learning (FL) by impacting the interoperability and consistency of models trained across diverse data sources, potentially compromising model performance and reliability.	Future research in healthcare could explore advanced collaborative strategies using Federated Learning (FL) across multiple institutions to enhance data sharing and analysis capabilities while maintaining data privacy and security. Investigating the integration of FL in precision medicine could lead to personalized healthcare solutions tailored to individual patient needs, improving treatment outcomes and patient care quality. Additionally, integrating FL with Explainable AI in healthcare settings could enhance the transparency and interpretability of AI models, enabling healthcare professionals to better understand and trust the decision-making process.
Federated	Federated learning	Collaborative	Communication	Future research in

Learning for Healthcare Informatics, by Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, Fei Wang [8]	(FL) tackles privacy concerns with techniques like differential privacy (DP). Secure multi-party computation (SMC) can even be combined with DP to further reduce privacy risks. FL frameworks like FATE utilize homomorphic encryption for secure computations. To handle diverse data in FL, multi-task learning (MTL) is employed, while deep Q-learning is an emerging approach for optimizing resource management.	Modeling: Enables efficient machine learning while ensuring legal compliance and data privacy.	efficiency is challenged in Federated Learning (FL) by the large number of clients and varying data distributions, leading to increased latency and potential bottlenecks in model updates. Efficient communication protocols are essential to mitigate these issues.	medical federated learning (FL) can advance by incorporating domain knowledge to personalize global models, thereby enhancing their accuracy and relevance in clinical settings. Efforts should focus on improving prediction models to better forecast future health conditions with precision. Optimizing communication is also crucial, requiring strategies to tackle efficiency challenges such as optimizing update sizes and exploring peer-to-peer learning approaches. These initiatives collectively aim to bolster the effectiveness and applicability of FL in healthcare, ensuring that advancements in technology translate into meaningful improvements in patient care and outcomes.
Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach, By- Akhil Vaid, Suraj K Jaladanki, Jie	This study evaluated federated learning for predicting COVID-19 mortality using electronic health records. Federated models using logistic regression and multilayer perceptrons generally	Improved Generalizability: Models trained on data from multiple institutions can be more robust and generalizable.	The study's focus on data from Mount Sinai Health System hospitals limits its scope, potentially affecting the generalizability of Federated Learning models to other	Expansion to other diseases. Integration with real-time data Enhanced privacy preserving techniques. Continuous model updates. Integration with clinical decision support system

Xu3, Shelly Teng [9]	surpassed local models trained on individual hospitals. While pooled models (combining all data) sometimes outperformed federated learning, this approach raises privacy concerns. These results highlight the promise of federated learning for improving healthcare predictions while protecting patient privacy.		regions with different patient demographics and healthcare practices.	
Split learning for health: Distributed deep learning without sharing raw patient, by Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, Ramesh Raskar [10]	Split learning and model partitioning are two techniques for training models on distributed data. Split learning divides the neural network itself, with different segments trained on separate servers. Model partitioning, on the other hand, keeps the model architecture but assigns different layers to be trained on client devices and a central server. Both approaches aim to improve privacy and efficiency in federated learning scenarios.	Overcoming Data Silos: In healthcare, where data is often fragmented across institutions, SplitNN allows collaboration without sharing sensitive information.	The complexity of implementing SplitNN (Split Neural Network) in Federated Learning demands specialized infrastructure and coordination among entities, which can increase operational costs and deployment challenges, hindering widespread adoption and scalability.	To enhance federated learning, privacy techniques, communication efficiency, scalability, automated model partitioning, and real-time on-device learning should be employed for data protection, seamless information exchange, and adaptability in dynamic environments.
Communication-efficient and differentially-private distributed SGD ,	The paper defines differential privacy and discusses its importance in the context of	Unified Solution: The proposed method effectively	In Federated Learning, balancing privacy with communication	Federated learning's future in healthcare is bright, but requires further development. Researchers are

By-Naman Agarwal, Ananda Theertha Suresh, Felix Yu, Sanjiv Kumar, H. Brendan McMahan [12]	distributed gradient descent. Differential privacy ensures that the addition or removal of a single data point does not significantly affect the output, thereby protecting individual data points. To reduce communication costs, the paper introduces the Binomial mechanism combined with quantization. Each client quantizes its data and adds noise drawn from a Binomial distribution before sending it to the server.	addresses both communication efficiency and differential privacy, which were previously treated separately.	efficiency involves a trade-off where enhancing differential privacy through noise addition to gradients may compromise model accuracy, presenting a challenge in achieving both robust privacy protection and optimal model performance simultaneously.	tackling privacy with stronger methods, optimizing algorithms for performance under strict privacy constraints, and exploring applications beyond medicine. Real-world testing will solidify its effectiveness, paving the way for secure and transformative healthcare advancements.
Differential Privacy-enabled Federated Learning for Sensitive Health Data By-Olivia Choudhury, Aris Gkoulalas-Divanis, Theodoros Salonidis, Issa Sylla [13]	Federated learning enables training a global machine learning model on distributed data without ever sharing the raw data itself. To further safeguard individual privacy, differential privacy can be incorporated as an extension to this framework.	Real-world Evaluation: The approach is evaluated using real-world health data, demonstrating its feasibility and effectiveness.	In healthcare, Federated Learning's assumption of numerous sites for effective implementation is impractical, restricting its real-world applicability due to the difficulty in aggregating sufficient data diversity and volume across a smaller number of sites, which can hinder model robustness and performance.	Federated learning's potential extends beyond its current applications. Researchers are exploring its use in other areas with similar challenges (non-smooth loss functions) and envisioning its impact on healthcare through improved policy making, early disease prevention, and personalized medicine.

5. CHALLENGES

Federated Learning has a number of difficulties. Furthermore vulnerable to assaults are federated learning models. Anybody involved in the FL process, a hacked central server, or exploited local devices inside the learning framework, can introduce the attacks. In the context of FL, attacks include backdoors, data poisoning, model poisoning, and membership inference attacks.

Different FL models have emerged to address these issues and reduce communication overhead. One-shot federated learning: The majority of FL aggregation methods available now are intended for synchronized device operation. However, model transfer and training happen asynchronously because of the heterogeneity of the systems and data. As such, scaling federated optimization in a synchronized fashion might not be possible [14]. Asynchronous environments can be used to implement federated learning, as discussed in works like [15, 16, 14]. Asynchronous Federated averaging techniques may accommodate more devices and enable updates to come in one at a time, in contrast to FedAvg (which operates in a synchronized way).

Blockchain in FL: To manage the asynchronous arrival of parameters from the devices, an aggregator is required to update the global model. This may prevent the FL models from being widely adopted. Devices may collaborate to learn without the need for a central aggregator since blockchain is a decentralized network. Works in the blockchain architecture that propose Federated Learning include [17, 18, 19, 20]. Figure 2 displays an example of architectural [18].

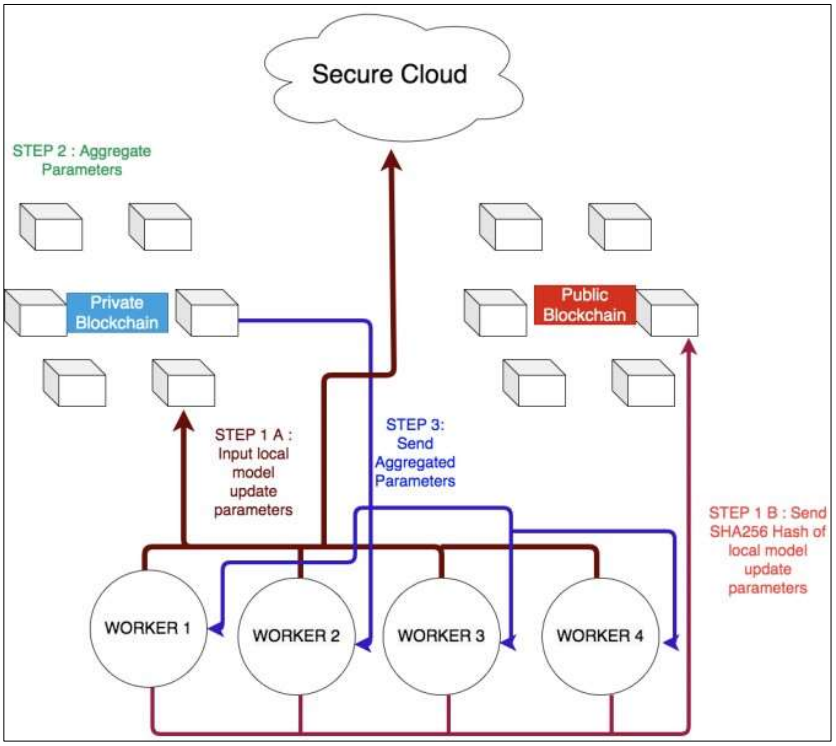

Figure 2: An example of a Federated Learning over Blockchain Architecture.

6. CONCLUSION

Federated Learning provides a safe environment for several devices to work together on machine learning without exchanging personal information. Numerous researchers were drawn to this, and a great deal of study is being done in this area. Federated learning has been used in a number of industries, including transportation and healthcare. FL frameworks are still vulnerable to a number of attacks even if they provide a stronger privacy guarantee than other ML frameworks. The deployment of defense mechanisms is further complicated by the scattered structure of the architecture. For example, when employing model poisoning defensive mechanisms, the gaussian noise introduced to the local models (for DP) may cause confusion for the aggregation methods and lead to the exclusion of the benign individuals. Thus, it would be interesting to investigate the following research question: Can techniques with minimal computational cost be used to construct a byzantine tolerant FL model that ensures user privacy.

REFERENCES

[1]. Brendan McMahan and Daniel Ramage. 2017. Federated learning: Collaborative machine learning without centralized training data. *Google Research Blog* 3 (2017).

[2]. AI pandemic engine. https://hai.stanford.edu/blog/pandemic-ai-engine-without-borders, 2020.

[3]. Tianlong Yu, Tian Li, Yuqiong Sun, Susanta Nanda, Virginia Smith, Vyas Sekar, and Srinivasan Seshan. 2020. Learning Context-Aware Policies from Multiple Smart Homes via Federated Multi-Task Learning. In *2020 IEEE/ACM Fifth Interna- tional Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 104–115.

[4]. Robin C Geyer, Tassilo Klein, and Moin Nabi. 2017. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557* (2017).

[5]. Minghong Fang, Xiaoyu Cao, Jinyuan Jia, and Neil Gong. 2020. Local model poisoning attacks to Byzantine-robust federated learning. In *29th USENIX Security Symposium (USENIX Security 20)*. 1605–1622.

[6]. Chengliang Zhang, Suyi Li, Junzhe Xia, Wei Wang, Feng Yan, and Yang Liu. 2020. Batchcrypt: Efficient homomorphic encryption for cross-silo federated learning. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. 493–506

[7]. A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics, By- K. Dasaradharami Reddy, Thippa Reddy Gadekallu

[8]. Federated Learning for Healthcare Informatics, by Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, Fei Wang

[9]. Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach, By-Akhil Vaid, Suraj K Jaladanki, Jie Xu3, Shelly Teng

[10]. Split learning for health: Distributed deep learning without sharing raw patient, By-Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, Ramesh Raskar

[11]. Simulation of denial of service (DoS) attack using matlab and Xilinx, I Mukhopadhayay, S Polle, P Naskar - IOSR Journal of Computer Engineering (IOSR-JCE), 2014

[12]. Communication-efficient and differentially-private distributed SGD , By-Naman Agarwal, Ananda Theertha Suresh, Felix Yu, Sanjiv Kumar, H. Brendan McMahan

[13]. Differential Privacy-enabled Federated Learning for Sensitive Health Data By-Olivia Choudhury, Aris Gkoulalas-Divanis, Theodoros Salonidis, Issa Sylla

[14]. Cong Xie, Sanmi Koyejo, and Indranil Gupta. 2019. Asynchronous federated optimization. *arXiv preprint arXiv:1903.03934* (2019).

[15]. Michael R Sprague, Amir Jalalirad, Marco Scavuzzo, Catalin Capota, Moritz Neun, Lyman Do, and Michael Kopp. 2018. Asynchronous federated learning for geospatial applications. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 21–28.

[16]. Marten van Dijk, Nhuong V Nguyen, Toan N Nguyen, Lam M Nguyen, Quoc Tran- Dinh, and Phuong Ha Nguyen. 2020. Asynchronous Federated Learning with Reduced Number of Rounds and with Differential Privacy from Less Aggregated Gaussian Noise. *arXiv preprint arXiv:2007.09208* (2020).

[17]. Xianglin Bao, Cheng Su, Yan Xiong, Wenchao Huang, and Yifei Hu. 2019. Flchain: A blockchain for auditable federated learning with trust and incentive. In *2019 5th International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, 151–159.

[18]. Harsh Bimal Desai, Mustafa Safa Ozdayi, and Murat Kantarcioglu. 2020. BlockFLA: Accountable Federated Learning via Hybrid Blockchain Architecture. *arXiv preprint arXiv:2010.07427* (2020).

[19]. Rajesh Kumar, Abdullah Aman Khan, Sinmin Zhang, WenYong Wang, Yousif Abuidris, Waqas Amin, and Jay Kumar. 2020. Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging. *arXiv preprint arXiv:2007.06537* (2020).

[20]. Paritosh Ramanan, Kiyoshi Nakayama, and Ratnesh Sharma. 2019. BAF- FLE: Blockchain based aggregator free federated learning. *arXiv preprint arXiv:1909.07452* (2019).