Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776869

SIGNIFICANCE OF FEDERATED LEARNING FRAMEWORK WITH DIFFERENTIAL PRIVACY PROTECTION FOR SMART HEALTHCARE

P. Karthiga[1], Dr Antony Selvadoss Thanamani[2], N. Balakumar[3], Dr A. Kanagaraj[4] S. Sathiyapriya[5], A. Shubha[6]

[1,3,5,6] Ph.D Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College Pollachi, Tamil Nadu - 642001, India. Email: karthigaprithivirajan@gmail.com

[2] Associate Professor & Head, Department of Computer Science, Nallamuthu Gounder Mahalingam College Pollachi, Tamil Nadu - 642001, India. Email: selvdoss@gmail.com

[4] Assistant Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College Pollachi, Tamil Nadu - 642001, India. Email: a.kanagaraj@gmail.com

ABSTRACT

Utilizing real-world health data for machine learning tasks necessitates addressing a number of practical issues, including distributed data silos, privacy concerns with creating a centralized database from person-specific sensitive data, resource constraints for transferring and integrating data from multiple sites, and the risk of a single point of failure. This invention established a privacy-preserving federated learning (PPFL) architecture capable of learning a global model using dispersed health data kept locally at many sites. Recent communication technology advancements have altered smart healthcare supported by artificial intelligence (AI). AI techniques have traditionally required centralized data gathering and processing, which may be impossible in realistic healthcare contexts due to the great scalability of modern healthcare networks and growing data privacy concerns.

Federated Learning (FL), an emerging distributed collaborative AI paradigm, is especially appealing for smart healthcare because it coordinates several clients to complete AI training without requiring raw data sharing. It is critical to develop FL models in a privacy-preserving manner, especially in the setting of healthcare, where patient data is extremely sensitive. This paper highlights the significance of FL with Differential Privacy (DP) for smart healthcare. Federated Learning is one of the most accepted solutions for training machine learning models since, unlike other strategies, it has no effect on system speed.

Keywords: Federated Learning, Differential Privacy, Smart Healthcare, Machine Learning, Artificial Intelligence.

1. INTRODUCTION

Federated Learning (FL) is a platform for smart healthcare systems that use wearables and other Internet of Things enabled devices. In recent years, the amount of privately owned medical data has grown rapidly due to wearable technology's accessibility, affordability, and increasing popularity. Wearable technology makes it possible to continuously gather and record data streams about physiological parameters, user mobility, and pertinent environmental

variables. Users can acquire personalized medical data sets that include details about their lifestyle and current state of health thanks to such data streams. Wearable technology is crucial for tracking health, safety, home rehabilitation progress, therapy effectiveness, early disease diagnosis, and other indicators of health condition.

By integrating home, mobile, and in-clinic health monitoring, wearable technologies provide continuous healthcare. Data analysis can be done in real-time for ongoing monitoring and alarm raising as necessary. The ability to monitor healthy people for little expense in order to spot anomalies and enable early diagnosis or prompt preventative measures is one of these technologies' main advantages. The continuous monitoring of human vital signs, activities, and other metrics during daily living at home, indoors, outdoors, or in healthcare institutions is made possible by wearable devices, the Internet of Things, and mobile health technology. The Smart Health Home and AAA medicine (Anywhere, Anytime, any environment) movements in healthcare and medicine are based on these technologies.

2. FEDERATED LEARNING

Federated Learning (FL) has emerged as a promising method for implementing low-cost smart healthcare applications with enhanced privacy protection. FL is a distributed AI technique that allows for the training of high-quality AI models by averaging local updates gathered from numerous health data clients without requiring direct access to the local data. Recent advancements in communication technologies and the Internet-of-Medical-Things have altered artificial intelligence-enabled smart healthcare (AI). AI techniques have traditionally required centralized data gathering and processing, which may be impossible in realistic healthcare contexts due to the great scalability of modern healthcare networks and growing data privacy concerns.

Federated learning, on the other hand, requires training many machine learning models on mobile devices and then combining their results into a single model that lives on a server. As a result, utilizing ground-truth data, a model is trained on devices, and only the learnt model is shared with a server. As a result, the user's data is used to create machine learning models while remaining private. In this case, federated learning benefits from the users' data without revealing any personal information about them. The raw data remains on the users' devices and is never transferred to a data center, but a model is built from this data and supplied to the server. Only local updates, such as model gradients, are required by the central server for AI training in the FL-based smart healthcare system, whereas local health data are stored at local medical sites and equipment. This would limit the dangers of sensitive user information being leaked to an external third party, offering a higher level of user privacy. Federated Learning will protect data privacy by training the ML model locally, without moving the data. In this setting, an ML model can still be changed or contextualized locally, which is more effective than relying on a single trained model.

3. FEDERATED LEARNING IN SMART HEALTHCARE SYSTEMS

Federated learning (FL), a new distributed collaborative artificial intelligence (AI) paradigm, is especially appealing for smart healthcare since it orchestrates numerous clients to provide

edge-based learning. FL achieves model learning without disclosing private data, unlike traditional machine learning (ML), which gathers all data on one site for centralized learning. Users of the FL frameworks can directly train models locally on their edge devices using their own private data sets without having to upload any data, and only submit parameter updates to the central server for model learning by aggregation. To protect user privacy, this approach solely requires the transfer of update parameters. FL ensures that multiple parties train ML models jointly, preventing the establishment of data islands. FL also contributes significantly to precision medicine by bridging the gap between home healthcare and traditional (hospital) healthcare.

Medical institutions have access to health informatics resources that allow for centralized data administration and compliance with regulatory, medical software, and medical device standards. The typical Smart Healthcare user, however, lacks the medical understanding and data necessary to govern data collection, management, and utilization. For the creation of models for disease prediction and health monitoring, pooling data from numerous wearable users is crucial. This also creates a critical mass for the engagement of health and data analytics professionals. However, data sharing for the creation of health monitoring models creates significant issues with safety, privacy, interoperability, and data integrity. Certain data-safety solutions, such anomaly detection and cognitive cyber-attack detection, can improve FL technology. Presently, the software environment is fragmented since many wearable devices communicate with the server using their own protocols.

Internet of Health advancements things make it possible to gather and parse health and medical data efficiently, but due to security and privacy concerns, centralized data collection is becoming more and more constrained by legal requirements. FL creates models without disclosing private data, in contrast to traditional machine learning (ML), which gathers all data on one site for centralized learning. Users of wearable devices can directly train models on their edge devices utilizing private data sets without having to submit data using the FL frameworks. FL uploads parameter updates to the main server for model learning through aggregation rather than sharing data. To protect user privacy, this technique simply entails transmitting updated settings. By collaborating on the joint training of ML models, FL avoids the problem of data islands.

4. MOTIVATIONS AND REQUIREMENTS OF USING FL IN SMART HEALTHCARE

The limitations and Key benefits of FL with current healthcare systems are as follows;

4.1 Limitations of Current Smart Healthcare Systems

Health information is subject to privacy breaches since standard AI-based approaches to achieving smart healthcare involve open data exchange with the cloud or data centers. In fact, attackers may obtain unauthorized access to AI training facilities to retrieve data, or outside parties, like cloud service providers, may seize control of the data and alter data patterns without the users' approval. In realistic healthcare systems, a single medical site's dataset could not be big enough to run an AI model, which would hinder it from properly learning from health data. This renders AI-based smart healthcare solutions ineffective and necessitates

manual data analysis, which results in significant delays in data processing. The data interchange between medical sites to enable data training is one potential solution, however due to institutional rules and rising user privacy concerns, it is challenging to acquire data from other sites to train the AI model.

The offloading of health data to the cloud for execution causes excessive network latency in the conventional AI-based smart healthcare systems, especially because medical data are frequently enormous in size. Also, the transfer of health data uses a lot of network capacity, which will probably result in network congestion as the number of devices rises. Medical device transmission power is also required throughout the offloading process, which presents new hardware and battery design problems.

4.2 Benefits of FL in Smart Healthcare

Only local updates, such as model gradients, are needed by the central server for the AI training in the FL-based smart healthcare system, whilst local health data are stored at neighborhood medical facilities and devices. This would increase user privacy by lowering the chances of the disclosure of sensitive user information to an outside third party. Thus, FL implementation is crucial for creating secure and reliable smart healthcare systems. FL is able to provide a reasonable trade-off between accuracy and utility together with privacy enhancement as compared to traditional centralized learning. Moreover, FL training preserves the generalizability of the model at the expense of a minimal accuracy loss. In exchange, FL's distributed learning function can improve the scalability of the smart healthcare system. FL can considerably save communication costs, like as latency and transmit power, consumed by raw data transfer by avoiding the offloading of large data volumes to the server. This is because model gradients typically have much smaller sizes relative to their actual datasets. FL also saves a significant amount of network capacity and reduces the likelihood of network congestion in large healthcare networks as a result.

4.3 Requirements

The central server, which aggregates local model gradients to construct the global model in each communication round, is one of FL's most significant entities. It has been established that model updates may still contain health user-related information such as data features and image resolution that can be reconstructed by the curious global server, despite the FL concept's ability to provide privacy protection by allowing users to keep their data at local sites during the training. So, user privacy may be compromised throughout the training, making FL vulnerable as a result and discouraging medical sites from participating in the cooperative training. Based on this reality, creating a trustworthy server for the coordination of data training and model aggregation is a vital necessity to guarantee reliable FL operations in smart healthcare. The dependable connectivity between local clients and the global server is a crucial prerequisite for FL-based smart healthcare. Due to potential external attacks, sending local model updates to the server could be dangerous. Since there aren't enough local updates to create a global model, an adversary may use data assaults to disrupt the communication channels that clients and servers have created in order to steal the updated information.

Their computational performance is a major concern in FL-based mobile smart healthcare as mobile medical devices participate in the federated data training. In order to acquire the ideal training performance, one must participate in the several communication rounds necessary to accomplish federated smart healthcare. In this context, some medical devices, such as small smart watches, might not be able to participate in the training over time due to their low computational power and energy resources. The FL concept becomes ineffective in smart healthcare without the participation of numerous devices, because the computations supplied by various devices are crucial to improving health data training. In order to create FL-based smart healthcare ecosystems, it is crucial to understand how to create computation-accelerated hardware for health devices. The availability of datasets at clients is required in FL-based smart healthcare to achieve the desired training performance. Based on its operating environment, each medical device or site must create its own dataset. For instance, a smartphone can capture patient mobility data while in use. Each participating client can also create its own data features for local training based on the dataset it has acquired using data-driven approaches like feature extraction. The non-independent and identically distributed problem, which may cause the FL training to be extremely divergent in the data training, is one of the main problems in this context when it comes to dataset preparation.

5. ADVANCED FL DESIGNS FOR SMART HEALTHCARE

Internet-of-Medical-Things and communication technology advancements in recent years have transformed smart healthcare (AI). Some of the advanced FL designs for smart healthcare are as follows;

5.1 Resource-aware FL

Using local model updates from IoMT devices in the uplink, communications with the aggregation server, and global model broadcasting in the downlink, collaborative FL models are created. Resource management is crucial in this aspect for enhancing the functionality of FL-enabled healthcare apps. By selecting a group of IoMT devices, a scheduling issue is a significant factor in reducing the overall training duration. The non-integer scheduling problem with enormous devices, however, is extremely difficult to solve. Optimizing shared computing and radio resources is a promising step towards resource-aware FL for smart healthcare. The success of AI-enabled numerous services and applications deployed simultaneously at mobile devices and the network edge inspired the joint compute and communication resource allocation paradigm for multiple FL services.

5.2 Secure FL

Since there are potential security attacks on FL systems such poisoning attacks, inference attacks, backdoor attacks, malicious servers, communication bottlenecks, and free-riding attacks, it is crucial to research secure solutions for FL-enabled healthcare applications. Throughout the past few years, various solution methods have been looked into in the context of smart healthcare. In the context of FL-enabled healthcare systems, a novel idea called reputation is presented to prevent unreliable updates from untrusted devices. Such dependable device selection is crucial in reducing the impact of various security assaults. The accuracy of

the global model may be compromised by altering the fall local model, for instance, if a malicious IoMT device injects poisonous data into its local data. It's also crucial to use a reliable device when local FL models are trained on noisy or poor-quality data. For secure federated smart healthcare, decentralized FL is a viable answer to the issue of unreliable parameter servers in centralized FL. Decentralized FL implementations frequently use on consensus, diffusion, and rumor methods.

5.3 Privacy-enhanced FL

FL has some privacy challenges, but its adoption can address many IoMT problems that are privacy-specific. For instance, using information from IoMT devices' local models, the global model automatically updates. By using construction assaults, the adversary can launch an attack and obtain user information. Different types of attacks are as follows;

- ***(i)*** **Information Leakage:** Information leakage is an application flaw when sensitive data, such as environment or user-specific information, or technical information about the online application, is exposed. An attacker may use sensitive data to take advantage of the target web application, its hosting network, or its users. Hence, if possible, efforts should be made to reduce or stop the leakage of sensitive information. Information leakage is most frequently caused by one or more of the following circumstances: Inappropriate application or server setups, a failure to remove HTML/Script comments holding sensitive information, or discrepancies in page responses for valid versus invalid input.
- ***(ii)*** **Poisoning attack:** The adversary in this type of assault adjusted the local model update parameters to lower the aggregator accuracy. There are two distinct poisoning attacks. for instance, model poisoning and data poisoning. The training data is affected in a data poisoning attack, whereas model parameters are attacked in a model poisoning assault.
- ***(iii)*** **Byzantine attack:** In this attack, the adversary or malicious node that participates in FL communicates the fictitious model parameters with the nearby nodes. Furthermore, the attacker may shorten the model's convergence time and accuracy by disseminating erroneous data.
- ***(iv)*** **Privacy data leakage attack:** In distributed FL, some training data information is continuously leaked during model updates shared with central servers by IoMT devices. The attacker can determine whether the end devices are dedicated to a specific task or not by starting the differential attack. In that case, the malicious user will obtain that data and use it for their own ends.
- ***(v)*** **Inference based attack:** Data mining is the main strategy used in this attack. The adversary uses data mining techniques during this attack to evaluate the data and extract some relevant information from it.

Differential privacy is the most well-known method that is frequently used to improve privacy preservation in numerous contexts. Many researchers started working on differential privacy based FL systems for IoMT and many other important areas as a result of this ability to prevent privacy leakage. Over the past ten years, a lot of scholars have studied the literature on IoT

security and privacy. It has been noted that solving security problems has received the majority of emphasis; nevertheless, in contemporary electronic health care systems, more focus needs to be placed on safeguarding end-user privacy. By merging cutting-edge digital technologies like IoT, high computing devices to store and process data, personal health records, and more, the health care system in the current day is converted into a new realm. It has been found that IoMT can considerably increase the efficiency of the healthcare system by ensuring patient safety through these cutting-edge counter measures.

5.4 Incentive-aware FL

All IoMT devices are required by conventional FL techniques to exchange local model updates with the aggregation server, although in practice this server is not always accessible. Because IoMT devices typically have constrained computer power, radio bandwidth, privacy concerns for user data, and server dependability, they are unwilling to share their models. Incentive-aware FL solutions are required to encourage the engagement of more FL users and enhance the effectiveness of FL-enabled healthcare scenarios. Incentive methods FL can be grouped according to a variety of criteria, including as the device's contribution to data, reputation, and resource distribution.

- Data quantity and data quality are two crucial measures used to assess data contribution. While the Shapely value is often used to measure data amount, data quantity actually refers to the size of local model updates and training samples.
- Device Reputation is a crucial measure for creating FL incentive systems. Reputation generally indicates how well FL users can supply high-quality data for model training and trustworthy local updates.
- Resource Allocation is a crucial stage of any incentive programme since it involves properly allocating computing and communication resources to FL users in order to boost FL performance.

5.5 Personalized FL

The standard FL confronts a number of difficulties when attempting to offer customized services to IoMT users. The global model only accurately represents the statistical traits of various IoMT devices, making it difficult to generate distinctive personal styles. The same body weight and height from different persons, for instance, may have different implications when learning to forecast the disease because of individual living situations and other external factors. Heterogeneous computing resources and network circumstances of various FL devices present another difficulty. The significance of individualized FL models has encouraged several advancements in personalized FL designs for smart healthcare in recent years.

6. FL APPLICATIONS IN HEALTHCARE

In a typical IoMT scenario, AI-based systems gather data from medical records, including disease diagnoses, medical pictures, clinical trials, drug discovery, and Electronic Health Records (EHR), among other things, for a variety of purposes. In this situation, there are serious privacy problems when exchanging EHR with other remote data centers or the cloud to

generate medical data. This shows that, especially in challenging healthcare contexts, removing or missing metadata such as patient information may not be sufficiently protecting privacy. The majority of AI systems rely on a central server because data must be supplied to analyze. FL suggests alternatives by combining more data with raised privacy awareness. Some of the applications where FL implemented are as follows;

(i) **EHR Data Analysis:** Since that healthcare data is usually fragmented and private, it might be difficult to acquire accurate results across populations when using EHR analysis. This makes it difficult to develop generalizable, efficient analytical methods that call for a range of "big data"-based learning. FL has a lot of potential for working with many healthcare data sources while protecting patient privacy. FL employs a centralized strategy in which a worldwide server creates a model using inputs from several local healthcare setups, preserving the sensitive data in the appropriate local organizations.

(ii) **Healthcare Monitoring:** Physical and mental health issues are on the rise among people as a result of COVID-19. Because of this, preserving our physical and mental health entails managing and overseeing our own everyday healthcare. The merging of the IoT with healthcare services to provide intelligent medical services has led to the development of the IoMT. Privacy and security concerns have made it harder for people to utilize it widely. The IoT-based healthcare systems have been advancing constantly to handle the expanding need for patient data transport. The usage of numerous health devices has been made possible by this expansion, including smart technology-based sensors that may track and evaluate a variety of aspects of a person's health and, in certain situations, serve as a trigger for potential health events. ML and Deep Learning (DL) algorithms are being used by the majority of interconnected healthcare systems to create judgments automatically for precise forecasts. While this has accelerated disease detection and improved disease diagnosis, there are numerous limitations that render the systems inaccurate and ineffective, such as a shortage of labelled data for training models.

Recently, there has been an increase in interest in creating intelligent systems for remote healthcare administration that shifts focus from hospitals to homes. By training a global model from scattered houses under the management of a server, such as a cloud server, FL can be used to facilitate in-home health monitoring while limiting data leakage by storing user data locally. By combining a class-balanced dataset with its own personal data and changing model gradients with the updated dataset in a way that the cloud and all houses update simultaneously, the IoMT device at each home can develop a customized model in this regard using convolutional neural networks (CNNs). This not only efficiently addresses difficulties with unbalanced and non-IID data but also enhances individualized forecasts.

(iii) Imaging in the Medical Field and COVID-19: Medical imaging is another area where FL could be used in an IoMT setting. It is used in conjunction with a few medical facilities for procedures including the segmentation and diagnosis of brain tumors using MRI, CT scans, and chest X-ray pictures. AI and computer vision assist in detecting the COVID-19 infection level. Due to patient privacy and protection concerns, the hospital does not permit the exchange of medical data without authorization. Such training data required a lot of labour to compile. As a result, employing DL approaches to track the spread of the COVID-19 infection will require insufficient data samples. The FL method can be used to address these kinds of issues. It can eliminate the issue of data silos and generate a shared model without getting local data.

7. DIFFERENTIAL PRIVACY

Differential privacy falls under the category of randomization-based schemes because it perturbs data. There have been alternative randomization techniques, however differential privacy, as stated, just gives a description from which mechanisms can be constructed. It can be achieved without using noise addition. Since that it is dependent on the logical relationship between facts, it can also be categorized as semantic anonymization. As the data analyst still needs to learn and the adversary is still regarded as the same person, it does not go as far as semantic security in the context of cryptography, which states that nothing is learned from a plaintext, but it still offers semantic guarantees. Differential privacy with the strongest properties is ε-differential privacy, where ε is the bound on privacy. In other words, serves as a boundary for the privacy loss's absolute value ε. The privacy loss is bounded by with at least probability 1-δ, and a relaxation of (ε, δ)-differential privacy permits privacy breaches to occur based off a minimal chance. The privacy guarantee is still valid even if the attacker is aware of every other database entry since it is unaffected by background knowledge. As a result, it defends against highly skilled attackers.

The characteristic of an algorithm known as DP enables for the retrieval of reasonably accurate responses to hypothetical questions from its encoded output, which is often a database. The need to protect the privacy of people whose private information is contained in a database while also being able to accurately learn about the entire population while using the database is the driving force behind DP. DP does not indicate a binary idea, such as whether or not an individual's data privacy is guaranteed. Instead, DP creates a rigorous standard for gauging privacy loss that enables comparison of various strategies. By setting a budget for privacy loss, DP will therefore strictly limit the potential harm to a person whose sensitive information is in the database.

Laplace mechanism: It is usually employed for preserving privacy in numeric queries $f: N^{|x|} \rightarrow R^k$ which map databases $x \in N^x$ to k real numbers. At this point, it is important to introduce a key parameter associated to the accuracy of such queries, namely the l_1 sensitivity:

$$\Delta f := \max_{\|x-y\|_1=1} \|f(x) - f(y)\|_1 \quad \text{---------- (1)}$$

Since the above definition must hold for every neighboring $x, y \in N^{|x|}$ it is also denoted as global sensitivity. This parameter sets the degree of uncertainty (i.e., noise) to be inserted in

the output to safeguard the privacy of a single individual since it quantifies the greatest magnitude of change in the output of f related to a single data element.

Moreover, we denote as Lap(b) the Laplace distribution with probability density function with scale b and centered at 0. Given any function f: $N^x \rightarrow R^k$, the Laplace mechanism can be defined as

$$M_L(x, f(.), \in) \coloneqq f(x) + (Y_1, \ldots . Y_k) \text{ ---------- (2)}$$

Where the components Y_i are IID drawn from the distribution Lap ($\Delta f/\in$). In other words, each component of the output of f is perturbed by Laplace noise according to the sensitivity of the function Δf. It can be shown that this is an $\in$-differentially private mechanism with $\in = \Delta f /b$.

Exponential mechanism: For circumstances when directly adding noise to the output function (as with the Laplace mechanism) will utterly sabotage the outcome, a general DP technique has been proposed. As the objective is to maximize utility while maintaining anonymity, the exponential technique serves as the foundation for inquiries with arbitrary usefulness. For a given arbitrary range R, the utility function N^x x R $\rightarrow$ R maps database/output pairs to utility values. We introduce the sensitivity of the utility function as

$$\Delta u \coloneqq \max_{r \in R} \max_{\|x-y\|_{1 \leq 1}} |u(x, r) - u(y, r) \text{ ---------- (3)}$$

Where the sensitivity of u with respect to the database is of importance, while it can be arbitrarily sensitive with respect to the range $r \in R$. The exponential mechanism $M_E(x, u, R)$ is defined as a randomized algorithm which picks as output an element of the range $r \in R$ with probability proportional to $\exp(\in u(x, r)/(2\Delta u))$. When normalized, the mechanism details a probability density function over the possible responses $r \in R$. Nevertheless, the resulting distribution can be rather complex and over an arbitrarily large domain, thus the implementation of such mechanism might not always be efficient. It can be shown that this is a $(2 \in \Delta u)$ - differentially private mechanism.

Gaussian mechanism: The addition of Gaussian noise to the results of a numerical query is a DP mechanism. Compared to the previously mentioned differentially private mechanisms, it has two significant advantages:

Common source noise: The additional Gaussian noise is identical to the one that arises normally while working with databases. Additive noise: It is simpler to statistically examine this DP mechanism since adding two Gaussian distributions results in a new Gaussian distribution. Instead of scaling the noise to the l1 sensitivity, as we previously did with the Laplacian mechanism, it is scaled to the l2 sensitivity:

$$\Delta_2(f) \coloneqq \max_{\|x-y\|_{1=1}} \|f(x) - f(y)\|_2 \text{ ---------- (4)}$$

To sum up, the main idea behind DP mechanisms is adding a certain amount of noise to the query output, while preserving the utility of the original data. Such noise is calibrated to the privacy parameters $(\in, \delta)$ and the sensitivity of the query function.

7.1 Differential Privacy in Healthcare and Medical Systems

Healthcare and medical systems are one of the most attractive uses of linking the physical and digital worlds. A lot of healthcare applications, including real-time health monitoring, fitness programmes, remote health monitoring, and senior care, are made possible by this connection, which has a lot of potential in CPSs. Another potential use for this connection is the delivery of medication and medical care from far-off locations or houses. Similar to this, big data storage

for health records and data analytics surveys for improved disease early diagnosis are both in the development stage. As a result, one of the fundamental components of CPSs is thought to be the healthcare and medical systems. By reducing wait times, costs, and enhancing quality of life, modern healthcare systems outperform conventional healthcare systems. Furthermore, by guaranteeing their most effective use, the modernization of these systems enables optimal allocation of finite resources.

7.2 Federated Learning with Adaptive Differential Privacy

In federated learning, which is a distributed method of developing machine learning (ML) models, data is handled locally, and only targeted model changes and metrics meant for immediate aggregation are shared with a server that manages training. This increases user privacy by enabling model training on signals that are readily available locally without disclosing raw data to servers. DP is a mathematical framework that places restrictions on a person's ability to affect a computation's result, such as the parameters of an ML model. This is done by producing a probability distribution over the output models by bounding the contribution of every particular user and adding noise throughout the training process. The DP has a parameter (ε) that measures how much the distribution might vary when one or more users' training examples are added or removed (the smaller the better). Federated training, where a trustworthy server manages the training process, enforces formal and significantly strong DP guarantees in a centralized manner. This guards against outside adversaries who might try to examine the model.

A method called Distributed Differential Privacy (DDP) provides DP guarantees for a server coordinating training that is trustworthy yet curious. With ML applications that use sensitive data, privacy concepts like data minimization and anonymization are crucial in addition to fundamentals like transparency and consent. The idea of data reduction is structurally included into federated learning systems. FL limits access to data at all stages, processes individual users' data as quickly as feasible (early aggregation), and promptly discards both acquired and processed data. FL only communicates the bare minimum updates for a specific model training task. Anonymization, which means that the final model should not memorize information specific to a particular individual's data, such as phone numbers, addresses, and credit card numbers, is another criterion that is crucial for models trained on user data. FL does not, however, immediately address this issue on its own. This anonymization principle can be properly quantified thanks to the mathematical idea of DP. The following figure 1 shows the Federated Learning Framework with Differential Privacy Protection for Smart Healthcare.

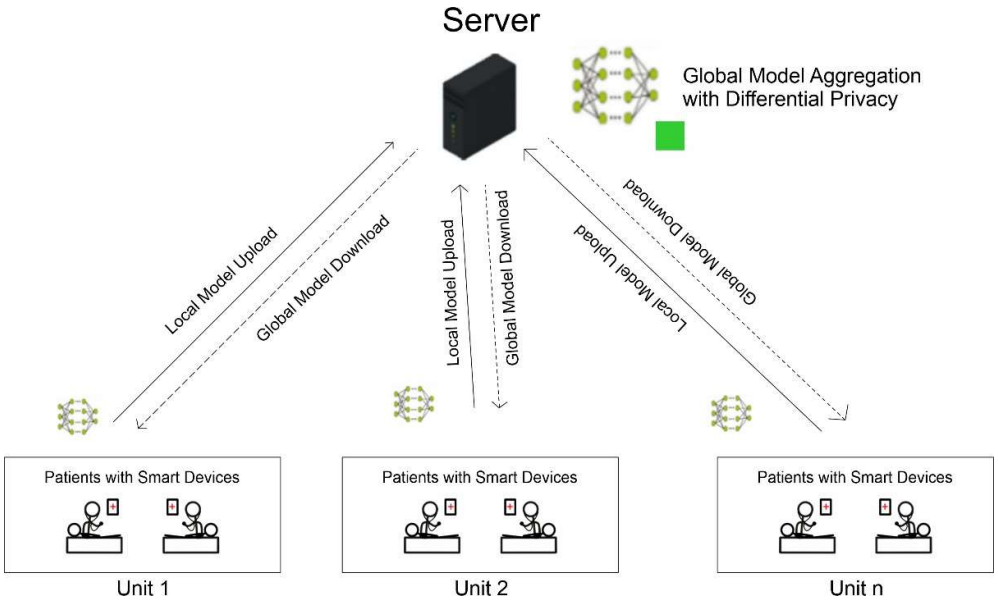

Figure 1: Federated Learning Framework with Differential Privacy Protection for Smart Healthcare

8. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

The main issues and future directions for FL-Healthcare research are presented in this section.

8.1 Communication Issues in FL-based Smart Healthcare

A key component of FL-enabled healthcare services is communication. Indeed, effective communication resource allocation plans can greatly enhance learning outcomes. When numerous IoMT devices are required to connect to the aggregation server for model updates in the uplink and model broadcasting in the downlink, this assumes increased significance. In this situation, the aggregation server can choose an appropriate group of IoMT devices using effective scheduling policies. The dynamic and quick changes in wireless channels, which have an impact on the dependability and quality of learning updates between IoMT devices and the aggregation server, represent another significant communication difficulty. Consider the impact of user abandonment along with more dependable design goals like outage probability and device availability as a potential remedy.

8.2 Quality of Federated Healthcare Training Data

The heterogeneity of computing capabilities and data characteristics across several hospital sites can significantly affect training quality. Designing incentive systems to encourage hospital/health organizations to submit accurate updates to the aggregate server and use high-quality data for training is a promising option in this scenario. Blockchain and game theory are two crucial resources for creating incentive structures. To enable FL entities like nurses, physicians, and patients to simply and appropriately adapt their actions, training requirements such as changes of data kinds, changes of learning rates, changes of training purpose-

classification, or regression, should be configured in a flexible manner. Due to the potential impact such modifications may have on the FL design and underlying learning models, adaptable FL techniques must be created.

8.3 Health Dataset Issues for Robust FL-based Health Data Analytics

Different clients may have distinct datasets, such as text, photos, audio, and time series, as well as diverse data contents, such as blood type, heart rate, facial images, and body temperature, in realistic healthcare settings. It is necessary to create new heterogeneous FL methods where the models of the cooperating parties may differ, but the central server can handle this heterogeneity through private ensemble learning.

8.4 FL-based Healthcare in Next-Generation Networks

Although 5G networks have not yet been widely adopted for commercial use around the globe, numerous efforts have been made to construct 6G wireless technologies. Many applications, including Industry 5.0, intelligent healthcare, the smart grid, and customized body area networks, are made possible by 6G. The introduction of numerous new technologies, including blockchain, compressive sensing, THz and visible light communications, 3D networking, quantum communication, and huge intelligence surfaces, coincides with the considerably tighter 6G criteria. Future research should focus on how to incorporate FL functionalities on future 5G/6G medical devices, how to use 6G devices, such as smart implants and wearables, for FL-based healthcare on a broad scale, and what new healthcare services 6G can enable. Future e-health services, for instance, will be improved by AI and FL capabilities, boosting patient quality of life and lowering hospitalization rates.

8.5 FL with Provable Privacy Guarantee

Despite FL's significant potential to safeguard user data privacy, there are many privacy concerns that need to be appropriately handled, particularly in situations related to smart healthcare given the high sensitivity of the health-related data. Membership inference attacks, unintentional information disclosure, and generative adversarial networks are three categories of FL privacy concerns. The attacker might make improper use of the global FL model to determine if a data sample is present in the FL health data collection. Also, when the patient's gadget transmits local model changes to the main server situated at the hospitals and healthcare facilities, the patient's information can be deduced. Using differential privacy, AI, and sophisticated cryptography approaches to provide privacy-preserving FL-healthcare system solutions appears promising.

8.6 Security Issues in FL-based Smart Healthcare

Many participants at the client side of FL-based smart healthcare systems may play the role of attackers and attempt to provide false or poisoned model updates in an effort to undermine the model aggregation. Moreover, a threat actor may taint data feature information during local data training or alter local updates during model transfer between local clients and the central server. An external adversary can use attacks on the server side to steal data from the aggregated global model, raising major privacy issues including data leaking. The solution to these security concerns is a significant barrier for FL-based smart healthcare systems. Consider other methods, such as employing differential privacy to safeguard training datasets against

data breach. Additionally, creating safe aggregation techniques is a viable way to enable key sharing between clients and the central server, provide a double-masking structure for encrypting local updates, and shield clients from data tampering and assault.

8.7 Non-iidness and Data Quality in FL-based Smart Healthcare

The non-iidness of the medical datasets, which may cause the FL training to diverge in the training, is a crucial issue that needs to be solved in order to obtain a good training performance in FL-based healthcare systems. A hospital might, for instance, have a higher distribution of a particular local disease than other hospitals in other locations. The label distributions in this instance vary between medical institutions, making it difficult for them to participate in the federated data training. Without tackling this non-iidness problem, the quality of the data training would considerably deteriorate or potentially diverge. In order to assure effective data training in FL-based smart healthcare, solutions to the non-iid problem need to be created, such as developing an additional subset of datasets to distribute equitably among clients. Using local batch normalization to correct the feature distributions at the client side before averaging the local models is another interesting method for implementing the feature shift among heterogeneous clients. In the FL-based smart healthcare sector, quantitative metrics are required to evaluate non-iid data, such as standard deviation, precision, and accuracy with respect to label/feature distribution skew and homogenous partitions.

9. CONCLUSION

The Internet-of-Medical-Things (IoMT) ecosystems, which collect real-time data from linked or implantable sensors, layered protocol stacks, lightweight communication frameworks, and end devices, have recently been adopted by healthcare companies. Healthcare analytics (HA) must be driven by IoMT in order to produce valuable data-based insights. Due to privacy laws, concerns have recently been raised about data exchange over IoMT and stored electronic health record (EHR) forms. As a result, the analytics model is judged to be erroneous with fewer data. Hence, a paradigm shift away from centralized learning towards distributed or edge-learning paradigms has begun in HA. Federated learning (FL) in distributed learning enables training on local data without requiring explicit data-sharing. FL's accuracy during the learning and updating processes is jeopardized by the significant statistical variability of its learning models, level of data partitioning, and fragmentation. The difficulties of large dispersed datasets, sparsification, and scalability issues have not yet been covered in recent surveys of FL in healthcare. The survey emphasizes possible FL integration in IoMT, FL aggregation policies, reference architecture, and the usage of distributed learning models to enable FL in IoMT ecosystems as a result of this gap. The distributed FL is practical for real-IoMT prototypes since it potentially outperforms the centralized FL techniques. The suggested survey could lead to outcomes that highlight important solutions and FL's ability to enable distributed, networked healthcare organizations.

References

[1] Anichur Rahman, Md. Sazzad Hossain et al., "Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues", Springer, Cluster Computing, 17 August 2022, DOI: https://doi.org/10.1007/s10586-022-03658-4.

[2] B. Murugeshwari, R. Sugumar, "Rule Based Privacy Preservation Method for Medical Data Sets", Middle-East Journal of Scientific Research 24 (8): 2640-2648, 2016, ISSN 1990-9233, IDOSI Publications, 2016, DOI: 10.5829/idosi.mejsr.2016.24.08.23876

[3] Fatima Alshehri, Ghulam Muhammad, "A Comprehensive Survey of the Internet of Things (IoT) and AI-Based Smart Healthcare", IEEE Access, Special Section on AI and IOT Convergence for Smart Health, Volume 9, 2021, DOI: 10.1109/ACCESS.2020.3047960.

[4] Georgios A. Kaissis, Marcus R. Makowski et al., "Secure, privacy-preserving and federated machine learning in medical imaging", Nature Machine Intelligence, VOL 2, June 2020, pp.305–311.

[5] Hui Cao, Shubo Liu et al., "IFed: A novel federated learning framework for local differential privacy in Power Internet of Things", International Journal of Distributed Sensor Networks, 2020, Vol. 16(5), DOI: 10.1177/1550147720919698.

[6] Jianzhe Zhao, Keming Mao et al., "Utility Optimization of Federated Learning with Differential Privacy", Hindawi, Discrete Dynamics in Nature and Society, Volume 2021, Article ID 3344862, 14 pages, https://doi.org/10.1155/2021/3344862

[7] Lina Ni, Chao Li et al., "Differential Private Preservation Multi-core DBScan Clustering for Network User Data", ScienceDirect, Elsevier, Procedia Computer Science 129 (2018) 257–262.

[8] Lina Ni, Peng Huang et al., "Federated Learning Model with Adaptive Differential Privacy Protection in Medical IoT", Hindawi, Wireless Communications and Mobile Computing, Volume 2021, Article ID 8967819, 14 pages, https://doi.org/10.1155/2021/8967819.

[9] Ons Aouedi, Alessio Sacco et al., "Handling Privacy-Sensitive Medical Data With Federated Learning: Challenges and Future Directions", IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, February 2023, DOI: 10.1109/JBHI.2022.3185673.

[10] R. Bharathi, T. Abirami, "Energy Aware Clustering with Medical Data Classification Model in IoT Environment", Computer Systems Science & Engineering, CSSE, 2023, vol.44, no.1, DOI: 10.32604/csse.2023.025336.

[11] Rongxin Qi, Sai Ji et al., "Security preservation in industrial medical CPS using Chebyshev map: An AI approach", Future Generation Computer Systems 122 (2021) 52–62, DOI: https://doi.org/10.1016/j.future.2021.03.008.

[12] Rui Hu, Yuanxiong Guo et al., "Personalized Federated Learning With Differential Privacy", IEEE Internet of Things Journal, vol. 7, no. 10, October 2020, pp.9530-9539.

[13] Tianqing Zhu, Ping Xiong et al., "Differentially private model publishing in cyber physical systems", Future Generation Computer Systems, https://doi.org/10.1016/j.future.2018.04.016, Elsevier.

[14] Tianqing Zhu, Ping Xiong et al., "Differentially private model publishing in cyber physical systems", Elsevier, DOI: https://doi.org/10.1016/j.future.2018.04.016, Future Generation Computer Systems.

[15] Wenshuo Wang, Xu Li et al., "A privacy preserving framework for federated learning in smart healthcare systems", Elsevier, Information Processing and Management 60 (2023) 103167.

[16] Xuefei Yin, Yanming Zhu et al., "A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions", ACM Computing Surveys, Vol. 54, No. 6, Article 131. Publication date: July 2021, DOI: https://doi.org/10.1145/3460427.

[17] Yu Wang, Zhenqi Huang et al., "Differential Privacy in Linear Distributed Control Systems: Entropy Minimizing Mechanisms and Performance Tradeoffs", IEEE Transactions on Control of Network Systems, vol. 4, no. 1, March 2017.

[18] Yuxia Chang, Chen Fang et al., "A Blockchain-Based Federated Learning Method for Smart Healthcare", Hindawi, Computational Intelligence and Neuroscience, Volume 2021, Article ID 4376418, 12 pages, https://doi.org/10.1155/2021/4376418.