Federated Learning for Privacy-Preserving Medical Diagnosis

[1]*Mamta Narwaria**, [2]*Shruti Jaiswal*
[1,2]*Department of Computer Science & Information Technology,*
Jaypee Institute of information technology, Noida, India

Abstract:
The field of medical imaging analysis has been substantially advanced because of computer technology, vision, and machine learning. The unprecedented success of modern machine learning techniques such as deep learning, could be because of the building and release of the grand-scale natural image databases, for instance, ImageNet, and Microsoft Common Objects in Context (MS COCO). Natural image analysis is fundamentally different from other areas, such as medical image analysis is still plagued by ‘‘small-sample-size’’ problem.
The intuitive and straight-forward solution for this small sample size problem is to pool images from different sites and build larger datasets to train high quality machine learning models. However, sharing medical imaging data between different sites/centers is intractable due to the strict privacy protection policies, such as Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR). Federated Learning (FL) presents a revolutionary solution to constructing medical diagnostic models while protecting patient privacy.
This research explores the application of FL in private medical diagnosis, including its architecture that includes secure aggregation mechanisms and privacy-preservation mechanisms like Differential Privacy, Homomorphic Encryption, and Secure Multi-Party Computation with its applications Disease Diagnosis while maintaining them private. Although it has its own strengths, FL faces some weaknesses related to communication overhead, data heterogeneity, and the possibility of adversarial attacks. some future areas of research work have been proposed, such as personalized federated learning, federated transfer learning, and use of blockchain. The Study show that Federated Learning serves as a viable solution in the scaling of diagnostic systems from the point of privacy perspectives in the application of artificial intelligence in the medical field.

Keywords: Federated learning, privacy- protection composition, medical diagnosis, safe multi-institutional cooperation, distributed machine learning, electronic health records.

1. Introduction:

The amalgamation of AI and machine learning in healthcare has seen great progress in diagnosis, treatment design, and disease prediction. Diagnostic models which are based on AI, which use deep learning, have reported very good results in the detection of many disease types which range from cancer and heart diseases to brain disorders and infectious diseases. However, it is a requirement to have large amounts of high-quality data which is usually held in many medical facilities. Trying to pool this data in a central location leads into serious privacy, security, and regulatory issues. In the past few years, the growth in data privacy concerns is seen, which is largely due to the introduction of strong data protection laws such as the GDPR in the EU and the HIPAA in the US. These laws put in place very strict rules around the sharing, storage, and processing of very personal info like health history. Also, traditional methods of training models which depend on bringing patient info together at one-point which are now impractical and legal gray areas. Also, increase in data breaches and unauthorized use of personal health info is the need for privacy preserving solutions in health care AI. Federated Learning which is a very promising solution to our privacy problems at the same time as it facilitates the building of very accurate and generalizable diagnostic models. It allows different medical institutions to collectively develop a universal model without requiring them to share private data due to its decentralized model of training. As opposed to typical machine learning which is a centralized approach, FL has a decentralized training which see multiple orgs train a global model without sending raw data to a central server. Instead, each participating institution trains the model with their own data and only model updates like weights and gradients go to a central aggregator which in turn puts the models together. Decentralization is what is guaranteeing that sensitive patient info stays secure within the walls of the institution thus preserving privacy and reducing the risk of data going out and being accessed by unauthorized people. The main reason is the growth in the use of FL in medical diagnosis is that it allows health care professionals to work together while at the same time protecting patient confidentiality. By using different data sets from many sources FL can improve accuracy in diagnosis especially for rare diseases which an individual institution may not have enough data on. Also, the fact that the models can be trained without breaking data privacy makes FL a great tool for improving diagnosis in resource poor settings and underserved areas. While FL has great promise it also brings with it a host of tech challenges. Data heterogeneity is an issue which arises when each institution has different data which plays a role in how the model performs. Also, the issue of communication overhead due to the frequent model updates which can play a role in scale up, especially for large models or

geographically dispersed participants. Also, FL systems are very much at risk to adversarial attacks which include data poisoning or model inversion which in turn play a role in compromising both privacy and model integrity. This study explored FL methods, report on their performance in other diagnostic tasks to achieve high diagnostic accuracy with experimental analysis.

2. Background:

[1] The article demonstrates the application of FL for MRI brain tumor image classification to overcome data privacy and collaboration challenges faced by conventional deep learning models, which compromises the secrecy of patients. The approach proposed here circumvents this by enabling different hospitals or research institutes to train a deep Convolutional Neural Network (CNN) model namely a VGG-based model in a manner where their raw data is not shared. This is especially valuable in the health industry where information is personal and regulating systems are stringent. Through the application of Federated Transfer Learning, the model leverages knowledge that is shared across institutions and generalizes. Experimental results indicate that the model performs with 98.4% accuracy, FL operates in medical imaging. The system maintains ownership of the data, is more secure and does not create data transfer legal entanglements. This research paves the way for scalable, privacy-protecting AI in diagnostic imaging and lays the groundwork for future research on secure medical image analysis with FL.

[2] Introducing a new Federated Learning (FL) framework for secure medical image sharing via chaos-based encryption. With more healthcare data being stored and processed in the cloud, patient data security is top priority. The study proposes a semi-synchronous FL system that combines Convolutional Neural Networks (CNNs) with the Henon Logistic Crossed Couple Map (HLCML), a chaos-based encryption method. Such a hybrid model fosters joint model learning and end-to-end encryption of private data such as hospital images on cloud servers. It supports delayed or partial model updates, to achieve robustness without compromising privacy.

[3] The survey presents a comprehensive taxonomy of FL methods, architecture, communication models, security protocols, and learning strategies with respect to healthcare applications. The survey explores disease prediction, medical image processing, and drug discovery. And highlights the challenges like heterogeneity of data, communication overhead, and personalization of models along with latest developments, points out gaps, and recommends directions for how to integrate FL into Healthcare 5.0 securely, scalably, and efficiently.

The EDP-AD system introduced in this paper is designed to identify Alzheimer's Disease (AD) at an early stage through the processing of audio data captured with the aid of IoT devices within geriatric patients' residences. Based on the privacy threats involved in sending confidential audio recordings to a central server, the authors combine Federated Learning (FL) and Differential Privacy (DP) for maintaining data confidentiality. The FL architecture enables local devices to learn a common model without revealing raw data, and DP puts noise in model updates, which also protects patient data. For more efficient communication and scalability, the system also employs a Top-k selection- based sparse mask update algorithm, decreasing the model update size. This technology conserves bandwidth without sacrificing model accuracy excessively. Experiments on real-world data sets demonstrate that EDP-AD achieves 84.48% accuracy in AD detection while ensuring strong privacy guarantees and reducing communication costs. The system proposed is a demonstration of how FL and DP can be combined effectively in real-world applications in the healthcare sector to enable secure, non-invasive, and efficient disease diagnosis through exploitation of smart home environments.

[4] In this paper, a privacy-protecting Deep Federated Learning (DFL) framework is proposed to identify new viral diseases from genome sequence data with confidentiality of patient data. Conventional molecular clinical diagnostics tend to involve centralized processing, where the risk of privacy violations is immense, particularly in pandemics or while monitoring outbreak diseases. To address this, the authors develop a decentralized framework in which genome information stays on devices locally, and model training is collectively performed with a version of LeNet deep learning architecture. The framework includes an automatic feature selection process to lighten computation and enhance learning efficiency. Crucially, it employs privacy-preserving methods like secure aggregation and differential privacy to protect sensitive genomic data during model updating. Experiments on real-world datasets indicate that the model achieves a 99.12% classification accuracy under i.i.d. data distribution among six clients. The model also performs better than benchmarks in metrics such as F1-score, ROC AUC, and Cohen's Kappa. The technique proves the viability of secure, distributed genomic surveillance with a robust mechanism for detecting and classifying unknown viral strains without violating data privacy.

[5] This work addresses an important limitation of federated learning (FL) in E-Health: the issue of spurious updates from client devices that do not significantly add to the global model. With the increase in edge devices gathering detailed electronic health record (EHR) data, FL provides model training in a privacy- preserving manner by sharing model gradients and not raw data. Nonetheless, certain clients might provide updates with

minimum utility or even damaging consequences, contributing to poor performance and convergence.

[6] This article presents a secure federated learning (FL) platform that combines chaos- based encryption with semi-synchronous updates to enable private medical image sharing and analysis. The platform uses Convolutional Neural Networks (CNNs) for image recognition and presents the Henon Logistic Crossed Couple Map (HLCML) for encrypting hospital This work addresses an important limitation of federated learning (FL) in E-Health: the issue of spurious updates from client devices that don't significantly add to the global model. With the increase in edge devices gathering detailed electronic health record (EHR) data, FL provides model training in a privacy-preserving manner by sharing model gradients and not raw data. Nonetheless, certain clients might provide updates with minimum utility or even damaging consequences, contributing to poor performance and convergence. In response, the authors introduce PFL-IU, a privacy-enhanced FL scheme that eliminates such meaningless updates. The system employs a secure communication-efficient aggregation protocol and a non-interactivity key generation algorithm to secure gradient data while in transit. Experimental findings indicate that PFL-IU enhances model convergence and predictive performance with the maintenance of client privacy. The framework is particularly useful in E-Health applications with highly heterogeneous client devices and data distributions to ensure that only meaningful and secure updates are added into the global model.

[7] This article presents DAFed, a Domain Adversarial Federated Learning model for learning from brain functional connectivity networks (FCNs) over distributed clinical sites. FCNs, constructed from resting-state functional MRI (rs-fMRI) scans, are important indicators of neurological disorders such as Alzheimer's disease and autism. However, limitations such as domain shift, data heterogeneity, and privacy issues hinder centralized training of models. DAFed addresses the above by integrating feature disentanglement, domain adversarial training, and contrastive learning in a federated environment. It separates extracted features into domain-invariant and domain-specific parts, allowing cross-site generalization while keeping client-specific patterns. The domain adversarial part allows for strong knowledge transfer across labeled and unlabeled sites, and contrastive learning enhances the global representation of invariant features. Experimental findings on multi-site datasets demonstrate that DAFed performs much better than conventional FL approaches in classification accuracy, particularly in heterogeneous and privacy-concerning settings. This framework emphasizes the strengths of domain-adaptive FL, such as making collaborative, privacy-protecting diagnosis of brain disorders possible worldwide.

3. Methodology

3.1 Federated Learning Framework for Medical Diagnosis

Federated learning (FL) is a new paradigm of machine learning. With FL, data in different medical institutions can be utilized in a decentralized fashion. In a conventional system, data is shared. However, with FL, only model updates are shared. FL is ideal for medical applications [8]. In medical applications, data security is of utmost importance. A conventional FL system has three main elements. They are training models locally, updating models, and aggregation.

I. **Local Training Process**

Each medical institution (e.g., hospitals, diagnostic labs) trains a local model on its dataset. This process ensures that patient data remains within the institution's secure environment.

- Data standardization: Each participant preprocesses its data using standardization protocols for compatibility across institutions. This includes normalization of features, image resizing and data cleaning.
- Model Initialization: A base model (e.g.,CNN for imaging) is shared with all clients to maintain architectural consistency.
- Weight updates generation: After training only the learned weights or gradients are shared. This maintains data locality and enhances privacy.

II. **Model update**

Once the local training is completed, each institution transmits its model update-not raw data for a central aggregator-or uses peer-to-peer strategies in a decentralized setup. These updates include:

- Gradients or weight updates from local model training.
- Alternative metadata such as loss price or local verification metrics. This mechanism ensures that patient learning is achieved without highlighting patient records or sensitive health information.

III. **Secure Aggregation: Hiding Individual Contributions**

The update received is collected to create a sophisticated global model. This model is revived to all customers for

the next training round[9]. The aggregation can be done through several ways:

- Federated average (Fedavg): A standard outlook that averages weight from all participating customers to the size of their dataset proportionally.
- Fedprose: A version of FedAVG which is responsible for non-IID data and client variability, improves convergence.
  - Safe aggregation:

Secure Aggregation indicated in Eq.1 ensures that the server can compute the sum of clients' model updates without learning any individual update. Secure Aggregation is robust to dropouts with advanced schemes [Bonawitz et al. (2017)]. It does not require encryption or decryption and this is vulnerable if fewer than a threshold of clients participate. The server should compute:

$$S = \sum_{i=1}^{N} w_i \qquad \text{.... Eq.1}$$

Where, w_i be the model update vector from client i,
N be the total number of clients.
Suppose N clients $\{C_1,\ldots,C_N\}$ each compute model updates Δw_i. Each client masks its update using pairwise shared random masks:

I. Each client i generates a mask (Eq.2):

$$m_i = \sum_{j=1}^{N} k_{i,j} \quad \text{.... Eq.2}$$

where $k_{i,j}=-k_{j,i}$ is a shared key between clients i and j.

II. Client i send (Eq.3):

$$\widetilde{w}_i = w_i + m_i \text{.... Eq.3}$$

III. Server aggregates (Eq.4) :

$$\sum_{i=1}^{N} \Delta \widetilde{w}_i = \sum_{i=1}^{N} \Delta w_i + \sum_{i=1}^{N} m_i \text{ .... Eq.4}$$

A cryptographic technique ensures that the individual model updates remain confidential during the aggregation process.The process is repeated over multiple rounds until the global model converges with optimal performance as shown in *Figure 1*.

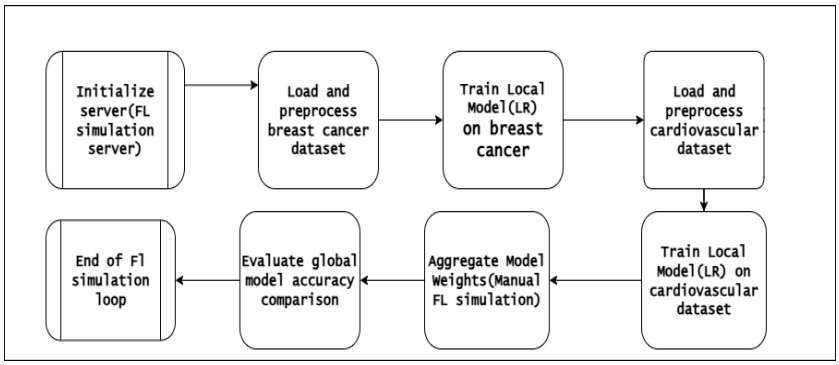

Figure 1. System design for the flow of process

Privacy-Preserving Technologies in Federated Learning

Privacy-protection technologies in federated learning, Although FL naturally provides privacy advantage, models updates can still leak sensitive data through projected attacks. To combat these risks, advanced privacy-protection methods are employed [10]. FL is one of the many privacy-protection techniques used in healthcare. Below is compared to other widely used methods:

Method	Privacy Mechanism	Advantages	Limitations
Federated Learning (FL)	Decentralized model training without data sharing	High privacy, regulatory compliance, improved generalization	Communication overhead, potential security threats
Anonymization	Removes personally identifiable information (PII)	Simple to implement, works well for structured data	May still allow re-identification through auxiliary data
Differential Privacy (DP)	Adds statistical noise to data or model updates	Strong mathematical privacy guarantee, prevents data leakage	Noise can reduce model accuracy
Homomorphic Encryption (HE)	Encrypts data for secure computation	Ensures complete data security, even during processing	Computationally expensive, slower than FL
Secure Multiparty Computation (SMPC)	Distributed computation without revealing data	High security for collaborative learning	Complex implementation, high resource usage

Table1: Comparison with other privacy methods

IV. FL Implementation in Healthcare Settings

To deploy FL in a clinical environment requires thoughtful adaptation to real -world obstacles, including data diversity, model complexity and communication infrastructure.

4.1 Model architecture: Depending on the type of medical data:

- CNNS (Conventional Neural Network) is used for medical image analysis (eg, detection of tumors in MRI/X-Ray).
- Time-series data such as RNNs (recurrent nerve networks) and LSTM ECGS are favorable for data.
- Transformers are gaining popularity in both imaging (vision transformer) and textual data (eg, clinical reports using Burt models).

4.2 Data characteristics

Medical datasets are naturally odd:

- Variability in imaging protocol: Various scanners and settings result in diverse input characteristics.
- Demographic discrepancies: Patient population varies in institutions, affects data distribution.
- Class imbalance: Rare diseases result in oblique dataset, which requires techniques such as overseas or class-weighting to ensure fair learning.

4.3 Communication protocol

FL can be deployed under various communication paradigms:

- Synchronous FL:All customers wait for others to complete local training before global update. It ensures synchronization but introduces the delay.
- Asynchronous FL: Customers update the global model independently as they complete their training, increasing the efficiency on the cost of potential incompatibility.
- Communication-efficient: strategies such as model prounting, perminuation, and federated dropouts help reduce bandwidth consumption and delay.

V. Results and Analysis:

The experimental analysis of Cancer Dataset expresses the highest Accuracy of 0.9737 and for the Cardiovascular Dataset, Accuracy is 0. 7238.The Confusion matrix for breast cancer disease as shown in Figure 2 & Confusion matrix for cardio disease is in Figure 3 respectively. Moreover, the Accuracy and precision for breast cancer & cardio diseases is shown Figure 4.

VI. Future Challenges:

While Federated Learning (FL) Prophet- Protection makes great promises to medical AI, several major areas should be addressed to ensure its practical deployment.

I. **Data inequality and non-IID distribution:** Medical data varies in hospitals due to imaging equipment, clinical protocols and differences in patient demographics. This give results in non- IID data, making the model

convergence difficult and reduces accuracy. Inconsistent labeling standard affects further model performance. Data asymmetries handle medical data in hospitals is highly diverse due to various devices, patient population and clinical protocols. Future FL model needs:

- Develop aggregation strategies for non- IID and unbalanced dataset.
- Integration of multimodal data such as imaging, EHR and genomics.
- Promote standardization in data formats and labeling.

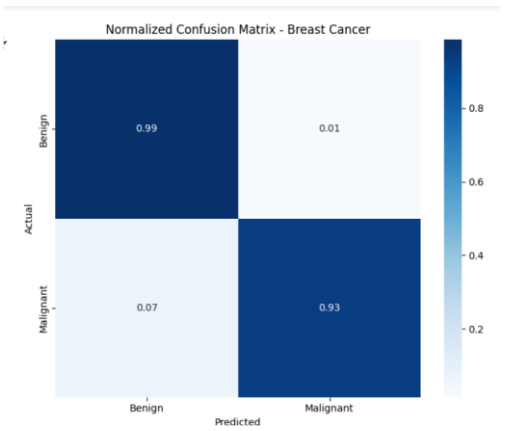

Figure 2. Confusion matrix for breast cancer disease

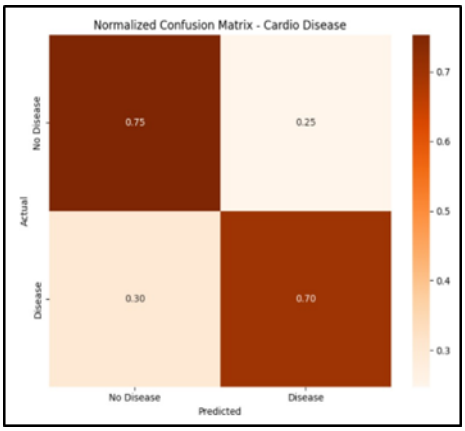

Figure 3. Confusion matrix for cardio disease

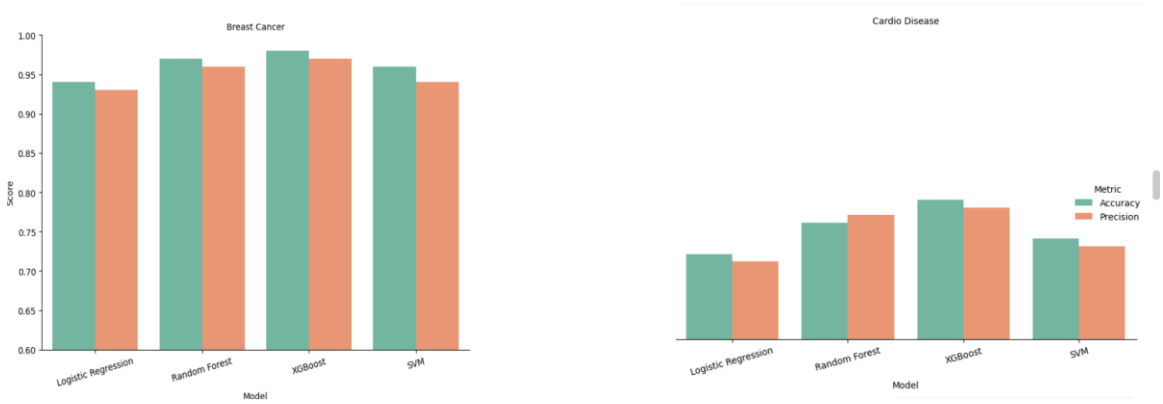

Figure 5. Accuracy and precision for breast cancer & cardio diseases

II. **High computational and communication overhead:**FL requires adequate computing resources and network bandwidth. Many hospitals, especially in low-resources settings, may lack the necessary infrastructure. Large models in synchronous FL may be delayed by updates and synchronization requirements and an increase in energy consumption.

III. **Safety and privacy risk FL:**Shields are unsafe for attacks such as leakage and model toxicity, where attackers can estimate personal information or corrupt global models. While confidentiality-protection techniques such as differential privacy and homomorphic encryption help, they also increase computational burden.

IV. **Difference and standardization issues:** Hospitals often use various electronic health records (EHRs) formats and preprosying methods, making the model challenging integration. The absence of universal FL framework works widely.

V. **Enhancing efficiency:** Enhance FL training is resource-intensive with high communication and energy costs. Solutions include: Compressing the model updates to reduce bandwidth and transfer to decentralized FL architecture using blockchain.

VI. **Strengthening privacy and security:** Although FL is secrecy-protection, there are risks.

VII. **Future Challenges:** Improve differential privacy methods for better accuracy-prone tradeoffs. Defense against shield leakage and model poisoning attacks. Creating a mildly secure aggregation protocol (eg, customized HE, SMPC).

VIII. **Navigate regulatory challenges:** Applying FL in areas requires compliance with laws such as HIPAA, GDPR and PDPB. Major directions include defining legal accountability for AI decisions and making FL models more interpretable and interpretable for physicians. Compliance with data protection laws such as Hipaa, GDPR, and India's PDPB is mandatory. FL should also navigate moral concerns such as the patient's consent and misuse of model by insurers or third party.

IX. **Participation and encouragement challenges:** Hospitals may lack incentive to participate in FL, especially if they do not feel any immediate benefit. Small institutions can contribute less due to resource limitations, which can lead to biased global models in favour of large hospitals**.** Adopt To succeed FL, hospitals must be motivated to participate. This includes:

- Making incentive models (eg, credit or token-based system).
- Ensuring small hospitals has the same effect in model training.

VII. Conclusion:

Federated Learning (FL) has emerged as a transformational approach to privacy- protection medical diagnosis, able to train hospitals and healthcare institutes to cooperate AI model collaborated, while ensuring that sensitive patient data remains localized. By eliminating the requirement of centralized data storage, FL HIPAA, GDPR, and PDPB align with global privacy regulations, making it a viable solution for real -world therapy AI applications. Additionally, FL takes advantage of diverse patient data from several sources, which improves the better model generalization and increased clinical accuracy. However, many challenges obstruct it widely adopting it. Data inequality remains an important issue, as hospitals use various electronic health records (EHRs) systems, imaging protocols, and clinical standards, resulting in non-IID data distribution.

References:

[1] Appasami, G. and Savarimuthu, N., 2025. Federated learning for secure medical MRI brain tumor image classification. *The European Physical Journal Special Topics*, pp.1-15.

[2] Roy, A., Mahanta, D.R. and Mahanta, L.B., 2025. A semi-synchronous federated learning framework with chaos-based encryption for enhanced security in medical image sharing. *Results in Engineering*, *25*, p.103886.

[3] Zhang, Y., Wang, L., Su, K.J., Zhang, A.,Zhu, H., Liu, X., Shen, H., Calhoun, V.D., Wang, Y. and Deng, H., 2025. A Privacy- Preserving Domain Adversarial Federated learning for multi-site brain functional connectivity analysis. *arXiv preprint arXiv:2502.01885*.

[4] Amin, M.S., Ahmad, S. and Loh, W.K., 2025. Federated learning for Healthcare 5.0: a comprehensive survey, taxonomy, challenges, and solutions. *Soft Computing*, pp.1-28.

[5] Zhang, W., Zhao, S. and Wang, H., 2025. A Privacy-Preserving System for Alzheimer's Disease Detection Based on Federated Learning. *Journal of Artificial Intelligence and Technology*.

[6] Mehedi, S.T., Abdulrazak, L.F., Ahmed, K.,Uddin, M.S., Bui, F.M., Chen, L., Moni, M.A. and Al-Zahrani, F.A., 2025. A privacy- preserving dependable deep federated learning model for identifying new infections from genome sequences. *Scientific Reports*, *15*(1), p.7291.

[7] Chen, H., Li, H., Xu, G., Zhang, Y. and Luo, X., 2020, June. Achieving privacy-preserving federated learning with irrelevant updates over e-health applications. In *ICC 2020-2020 IEEE international conference on communications (ICC)* (pp. 1-6). IEEE.

[8] Wei, W., Jammine, A. and Nait-Abdesselam, F., 2024, October. Enhancing Privacy Protection for Federated Learning with Distributed Differential Privacy. In *2024 20th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (pp. 408-413). IEEE.

[9] Zhang, Z., Wu, L., He, D., Wang, Q., Wu, D., Shi, X. and Ma, C., 2022. G-VCFL:Grouped verifiable chained privacy-preserving federated learning. *IEEE Transactions on network and service management*, *19*(4), pp.4219-4231.

[10] Amritanjali and Gupta, R., 2025, January. Federated Learning for Privacy Preserving Intelligent Healthcare Application to Breast Cancer Detection.