Differential Privacy-Enhanced Federated Learning in Medical Data Environments

Kanhaiya Jee Jha[1], **Dr. Gaurav Kumar Ameta**[2], **Esan Panchal**[3], **Keyurbhai A. Jani**[4]
[1]Swarrnim Startup & Innovation University, Gandhinagar, India, Kanhaiya.jeejha19@gmail.com, ORCID ID: 0009-0007-2804-1778
[2]Parul University, Vadodara, India, gauravameta1@gmail.com, ORCID ID: 0000-0002-7463-2583
[3,4]Information Technology Department, Government Polytechnic Gandhinagar, Gujarat Technological University, Ahmedabad - 382424, Gujarat, India.
[3]esan.gpg@gmail.com, ORCID ID: 0000-0002-1667-0864
[4]keyur.soft@gmail.com, ORCID ID: 0000-0002-6050-9365

Abstract- *In today's data-driven world, safeguarding the privacy and security of sensitive medical information particularly disease-related data has become a pressing concern. This research investigates the application of federated learning, differential privacy, and federated averaging to enable secure and private analysis of healthcare data. A novel framework is proposed that integrates these advanced privacy-preserving techniques to ensure individual data remains confidential while allowing collaborative analytics among various healthcare institutions. Through simulations and experimental evaluations, the framework's ability to protect patient privacy without compromising data utility is assessed. The results highlight the potential of this approach to support secure data sharing and analysis in modern healthcare environments, contributing to the advancement of privacy-centric health data solutions.*

Keywords- *Differential Privacy, Noise-based Privacy Protection, Privacy-Preserving Data Perturbation, FedAvg, Distributed Model Aggregation, Federated Learning, Decentralized Machine Learning, Collaborative Learning without Data Sharing, Edge-Based Model Training, Health Data Privacy**,** Confidentiality of Medical Records, Protection of Patient Information, Secure Handling of Health Data. Privacy Protection**,** Data Confidentiality Measures, Privacy Preservation Techniques, Information Security Controls, Sensitive Data Protection.*

1. INTRODUCTION

Differential privacy is a principle and methodology used in privacy-aware data analysis and statistical processing. It aims to solve the problem of gaining meaningful knowledge from sensitive datasets without compromising the privacy of individuals. The core objective of differential privacy is to guarantee that the results of any analysis remain virtually unchanged whether or not any one person's data is included, thereby ensuring that no single individual's information can be inferred from the output.

It is a mathematical way to protect people's personal information while still allowing useful data analysis. It helps ensure that no one can figure out details about a specific person, even if their data is included in a large dataset. This is especially important today, as huge amounts of personal data are being collected and used for things like research, creating policies, and training machine learning models[1].

The main idea behind **differential privacy** is to introduce carefully measured randomness—known as *noise*—to the data or the output of computations. This makes it difficult for anyone analyzing the results to determine whether a specific person's data was used. The added noise protects individual privacy while still preserving the overall statistical patterns in the dataset, reducing the risk of exposing sensitive personal information [2].

Core features and foundational principles of differential privacy are:

- **Privacy Assurance:** Differential privacy offers a measurable way to assess privacy protection, often represented by the parameter ε (epsilon). The smaller the ε, the stronger the privacy protection for individuals in the dataset.

- **Use of Randomization:** To implement differential privacy, data is deliberately modified using methods such as adding statistical noise, randomizing inputs, or altering data collection procedures. These techniques help obscure individual contributions.
- **Aggregate-Level Analysis**: Working with summarized or grouped data—like averages or totals—helps safeguard personal information by reducing the chance of identifying any one person's data.
- **Mathematical Foundation**: Differential privacy is supported by a precise mathematical structure, enabling researchers and data scientists to formally define, assess, and ensure the privacy guarantees offered by a specific approach.

Differential privacy has become increasingly popular, particularly in fields where protecting personal information is critical—such as healthcare, financial services, and public sector data analysis. As technology evolves, the role of differential privacy in maintaining a balance between useful data analysis and individual privacy is becoming more essential. Experts and developers are actively working on new methods and real-world applications to improve the efficiency and usability of differential privacy in practical environments [3].

Federated Learning is a method in machine learning that allows model training to occur across multiple decentralized devices or servers, with each device keeping its data locally. Instead of sharing raw data, devices work together to train a shared global model by only exchanging updates or learned patterns. This technique is especially useful in sensitive fields like healthcare, finance, and IoT, where maintaining data privacy and security is essential[4].

A major problem that **Federated Learning** aims to solve is the difficulty of centralizing data for training due to privacy issues, legal limitations, or the massive amount of data spread across different devices [3]. Rather than transferring raw data to a central server, Federated Learning enables each device to process and update the model locally. Only the learned updates (not the original data) are shared with a central server or peer devices. This ensures that sensitive data stays on the user's device, significantly enhancing privacy and data security [3].

Federated Averaging (FedAvg) is a commonly used algorithm in **Federated Learning** that combines model updates from multiple devices to create a shared global model. The general process includes the following steps [5]:

1. **Initialization:** The central server begins by creating and distributing an initial version of the global model to all participating devices.
2. **Local Training:** Each client or device uses its own local dataset to train the model independently. This training can involve several internal iterations to enhance performance.
3. **Model Update Calculation:** Once local training is completed, each device calculates the changes (or updates) made to the model during training compared to the original global model.
4. **Aggregation:** The central server collects the model updates from all devices and combines them—typically by averaging the updates—to form a single, unified update.
5. **Global Model Refinement:** The aggregated update is then used to refine the global model on the central server.
6. **Repetition of Rounds: Steps 2** through **5** are repeated across several training rounds, gradually improving the global model while keeping the raw data on local devices, thus preserving privacy.

Federated Averaging enables collaborative model training without compromising data privacy. By combining and averaging updates from many devices, it reduces the influence of noisy or abnormal updates, helping to produce a more stable and accurate global model [5].

Federated Learning, along with techniques such as **Federated Averaging**, has attracted growing interest across multiple fields. It provides a privacy-focused alternative to conventional centralized machine learning by enabling model training without direct access to raw data.

2. LITERATURE REVIEW

Federated Learning has shown strong potential for managing heterogeneous medical datasets in practical settings by leveraging clusters of machines for distributed processing. In this study, experiments were carried out using **CloudLab**, a specialized platform designed for research in distributed systems and networking. Multiple deep learning architectures and federated optimization methods were evaluated. Among the models tested, **Inception-v3** and **EfficientNetB0** consistently delivered the best results, achieving high accuracy on test datasets. In terms of optimization strategies, **FedAvg** outperformed others, with **FedAvgM** ranking as the second most effective. These findings highlight both the capability of federated learning in healthcare applications and the importance of selecting suitable models and optimization strategies to maximize performance[6].

To protect patient privacy, a novel approach combines **homomorphic encryption** with **federated learning** to create a secure diabetes prediction system. The experimental findings demonstrate that this method effectively breaks down data silos between hospitals, enabling the collection of patient data from multiple healthcare providers without compromising privacy. This practical and forward-thinking solution is especially relevant in today's data-sensitive environment, offering promising advancements for diabetes diagnosis and care. Additionally, it paves the way for new approaches to multi-party data integration, with potential applications across various sectors in the future[7].

With the rise in personal data privacy breaches, there is an increasing demand for methods that prioritize the protection of user information. To address this challenge, a federated learning algorithm has been proposed to predict breast cancer using data sourced from multiple hospitals. This technique safeguards patient privacy by enabling hospitals to collaboratively train machine learning models without sharing sensitive data with a central server. To evaluate its effectiveness, the federated approach was compared to traditional centralized methods. The findings revealed that the federated model delivered accuracy comparable to standard techniques. While the approach offers significant privacy benefits, it also presents certain limitations, which are thoroughly examined in this paper along with a detailed introduction to the concept of federated learning[8].

The integration of **Federated Learning (FL)** with **Software-Defined Networking (SDN)** presents a powerful solution for effective malware detection and mitigation, aiming to build a secure, automated, and privacy-aware network infrastructure within the healthcare sector. As hospitals increasingly rely on Information and Communication Technologies (ICTs), the continuous emergence of sophisticated malware attacks has created persistent uncertainty in the industry. Despite rapid advancements in medical technologies and device interconnectivity, many healthcare providers and patients have yet to fully embrace or understand these opportunities—leading to fragmented progress.

This research proposes a federated learning framework involving four geographically distributed hospital networks, enabling collaborative model training while preserving data privacy. The system utilizes **logistic regression with cross-entropy loss** for malware detection, ensuring high accuracy in identifying threats. SDN complements this framework by enabling dynamic network management and enforcing security policies, especially during the initial development and mitigation stages[11].

The experimental results highlight the model's effectiveness in maintaining accuracy without compromising patient privacy. This approach challenges the reliance on traditional centralized systems, which, while functional, often fail to provide adequate privacy safeguards in sensitive healthcare environments[11].

Federated Learning (FL) is a decentralized machine learning paradigm that enables devices to collaboratively train a global model without sharing raw data. This study builds upon the foundational **Federated Averaging (FedAvg)** algorithm by integrating principles from **consensus theory**. Unlike traditional FL methods that rely on a central coordinating server, the proposed method—called **FedLCon**—operates without one, thereby eliminating the risk of a single point of failure and reducing the need for mutual trust among clients. Furthermore, the consensus mechanism is also applied to the **Adaptive Federated Learning (AdaFed)** algorithm, an enhanced version of FedAvg that includes adaptive model averaging. The effectiveness of these approaches is evaluated through performance comparisons in a real-world use case: **COVID-19 detection**[12].

Federated Learning (FL) enables multiple participants to collaboratively train a global predictive model without exposing their private data. However, even with privacy-preserving mechanisms in place to protect local updates, a major challenge arises when users contribute low-quality or inconsistent updates. These irregular users can hinder model convergence and degrade overall performance. While some recent studies attempt to address both privacy concerns and the impact of irregular users, existing solutions often struggle with limited accuracy and efficiency. This is largely due to the overhead of complex cryptographic techniques and ineffective strategies for filtering out unreliable participants. To overcome these limitations, we propose **SAP-IU**—a novel and efficient federated learning framework that simultaneously ensures **privacy protection** and **irregular user mitigation**. At the core of our method is **TrustIU**, an innovative algorithm that assigns weights to users based on **cosine similarity**, allowing the global model to prioritize contributions from users with high-quality data. To further enhance privacy, we introduce a **secure weighted aggregation protocol** that protects sensitive information, including local model updates and data quality indicators. Additionally, our approach is designed to remain effective even when users drop out during training. Extensive experimental results demonstrate that **SAP-IU** surpasses existing methods in both training accuracy and computational efficiency [13].

Modern computer-aided diagnosis systems that leverage deep learning have become essential tools in medical imaging. As collaborative disease diagnosis across multiple healthcare institutions gains momentum, it also presents significant challenges—particularly the high annotation workload required from medical experts and the privacy and generalization limitations of centralized learning systems. To address these issues, we propose two novel **federated active learning** strategies: **Labeling Efficient Federated Active Learning (LEFAL)** and **Training Efficient Federated Active Learning (TEFAL)**, designed to support multi-institutional disease diagnosis.

LEFAL utilizes a **task-independent hybrid sampling strategy** that balances data uncertainty and diversity to enhance labeling efficiency. In contrast, **TEFAL** focuses on improving training efficiency by evaluating **client informativeness** through a discriminator-based mechanism. Experimental results demonstrate the effectiveness of both methods: on the **Hyper-Kvasir dataset** for gastrointestinal disease diagnosis, LEFAL achieves **95% segmentation performance** using only **65% of labeled data**. On the **CC-CCII dataset** for COVID-19 classification, TEFAL reaches an **accuracy of 0.90** and an **F1-score of 0.95** within just **50 training iterations**.

Overall, these federated active learning approaches outperform existing state-of-the-art techniques in both segmentation and classification tasks, offering a scalable and privacy-preserving solution for collaborative medical diagnosis across distributed healthcare centers [14].

Historical medical records are vital for improving healthcare by enabling intelligent diagnosis and disease prediction. Traditional smart health systems often rely on collecting data from multiple hospitals and labs, using machine learning algorithms for disease forecasting. However, these systems face limitations—particularly due to fragmented patient data, as individuals often consult different specialists across various healthcare facilities during treatment [8].

To overcome this challenge, we propose a **secure and intelligent federated learning framework** for health diagnosis that incorporates a **blockchain-based incentive mechanism** and a **non-fungible token (NFT)–powered data marketplace**. NFTs are used to define clear ownership and access rights for patient medical data, while the marketplace manages controlled access to historical records. The incentive mechanism rewards or penalizes patients based on key factors such as data quality, relevance, and upload frequency, encouraging meaningful contributions to the federated learning process. For model aggregation, we utilize the **Polyak-averaging** technique to merge local models into a unified global model. Extensive evaluations show that this decentralized framework delivers predictive performance comparable to centralized models, while also offering enhanced data security and access to high-quality data. The results emphasize the effectiveness of the blockchain-driven incentive system in encouraging patient participation and elevating the overall quality of the global health model [15].

The **Industrial Internet of Things (IIoT)** is a key component of **Industry 4.0**, where smart technologies and automation are transforming industrial operations. When combined with **machine learning**, IIoT enables the creation of intelligent and efficient industrial systems. However, a major concern arises from the use of sensitive data to train machine learning models. Sharing this data can lead to potential privacy breaches, posing serious risks to data security within IIoT environments. To address this challenge, we propose a **privacy-preserving data aggregation scheme** called **FLPDA**, built upon the **federated learning** paradigm. FLPDA allows data aggregation while safeguarding individual model updates, thereby preventing reverse-engineering attacks from centralized industrial administration centers. In each aggregation cycle, the **PBFT (Practical Byzantine Fault Tolerance)** consensus algorithm is employed to select an IIoT device within the region as the aggregator node. To enhance fault tolerance and secure data sharing, we integrate the **Paillier cryptosystem** with **secret sharing techniques**. Their comprehensive security and performance evaluations demonstrate that FLPDA effectively protects data privacy and resists various attack scenarios. Moreover, it achieves lower **communication**, **computation**, and **storage overhead** compared to current approaches. Simply put, FLPDA keeps sensitive industrial data private and secure—while being more efficient and scalable than existing solutions[18].

When the Internet of Things (IoT) is deeply integrated with healthcare, it forms the Internet of Medical Things(IoMT). In this ecosystem, doctors can diagnose and treat diseases using patient data collected from mobile and wearable devices, analyzed through AI-powered systems. However, conventional AI models may unintentionally expose sensitive patient information, raising serious privacy concerns. To address this issue, authors propose a privacy-enhanced approach using Federated Learning (FL) for IoMT-based disease diagnosis. FL enables multiple healthcare providers to collaboratively train a shared model without exchanging raw data. While FL improves privacy, it remains vulnerable to inference attacks, where malicious actors attempt to extract sensitive information from shared model updates.[19]

Our solution introduces a two-fold defense mechanism: First, we reconstruct medical data using a variational autoencoder (VAE) to transform it into a more privacy-resilient format. Next, we apply differential privacy by adding calibrated noise to protect against potential inference attacks. These privacy-preserved representations are then used to train local diagnostic models, ensuring that patient data remains confidential throughout the process. To further motivate participation in the FL process, we propose a reward-based incentive mechanism.

We evaluated our method using the MIT-BIH arrhythmia database, and the results confirmed that our approach effectively reduces the risk of patient data reconstruction while maintaining high accuracy in heart disease diagnosis. In summary, our framework strengthens privacy in medical AI systems without sacrificing diagnostic performance. [19]

3. Motivation For Proposal

As data breaches and privacy concerns continue to rise, especially regarding the handling of sensitive information, there is an urgent need for secure and privacy-preserving methods of data analysis and sharing. Focusing on **privacy-enhancing techniques**—particularly in the context of **disease data**—offers a timely and highly relevant area of exploration.

Investigating the integration of **differential privacy**, **federated learning**, and **federated averaging** provides a promising avenue for advancing data privacy in healthcare. This intersection could yield novel insights, methodologies, and solutions that not only protect individual privacy but also enhance the effectiveness of collaborative data analysis across institutions.

By combining these techniques, researchers can enable secure collaboration and data-driven discovery without compromising personal health information. This strategy not only tackles today's pressing privacy challenges but also lays the groundwork for broader innovation in privacy-preserving technologies, extending far beyond healthcare into various data-sensitive domains.

4. Dataset

The dataset comprises 4,920 records with 134 features. The data types of these features include one float64, 132 int64, and one object type. A few of these features are listed below [5]:

Table-1 Few Features

Itching	high_fever	Phlegm
skin_rash	sunken_eyes	throat_irritation
nodal_skin_eruptions	Breathlessness	redness_of_eyes
continuous_sneezing	Sweating	sinus_pressure
Shivering	Dehydration	runny_nose
Chills	Indigestion	Congestion
joint_pain	Headache	chest_pain
stomach_pain	yellowish_skin	weakness_in_limbs
Acidity	dark_urine	fast_heart_rate
ulcers_on_tongue	Nausea	pain_during_bowel_movements
muscle_wasting	loss_of_appetite	pain_in_anal_region
Vomiting	pain_behind_the_eyes	bloody_stool
burning_micturition	back_pain	irritation_in_anus

The target variable **'Prognosis'** includes the following class labels:
['Fungal infection', 'Allergy', 'GERD', 'Chronic cholestasis', 'Drug Reaction', 'Peptic ulcer disease', 'AIDS', 'Diabetes', 'Gastroenteritis', 'Bronchial Asthma', 'Hypertension', 'Migraine', 'Cervical spondylitis', 'Paralysis (brain hemorrhage)', 'Jaundice', 'Malaria', 'Chicken pox', 'Dengue', 'Typhoid', 'Hepatitis A', 'Hepatitis B', 'Hepatitis C', 'Hepatitis D', 'Hepatitis E', 'Alcoholic hepatitis', 'Tuberculosis', 'Common Cold', 'Pneumonia', 'Dimorphic hemorrhoids (piles)', 'Heart attack', 'Varicose veins', 'Hypothyroidism', 'Hyperthyroidism', 'Hypoglycemia', 'Osteoarthritis', 'Arthritis', '(vertigo) Paroxysmal Positional Vertigo', 'Acne', 'Urinary tract infection', 'Psoriasis', 'Impetigo'].

Figure 1 illustrates the Feature wise Mean and Standard Deviation, while Figure 2 displays the skewness of each feature and Figure 3 show Correlation Heatmap.

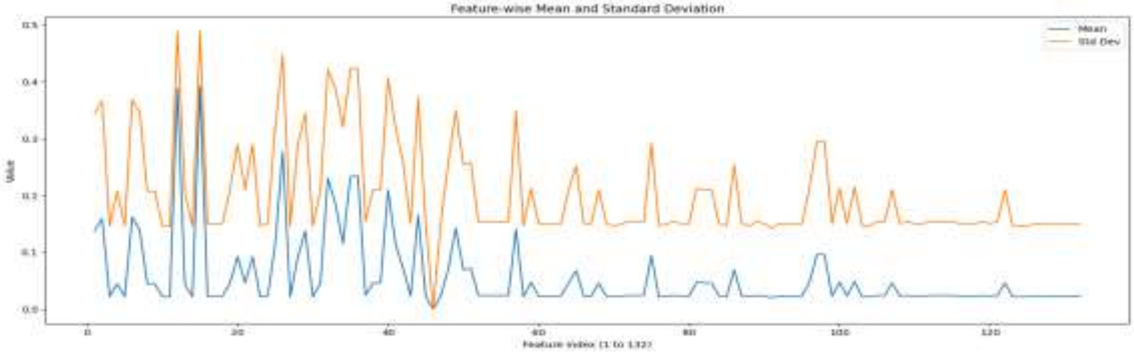


Fig-1 Feature wise Mean and Standard Deviation

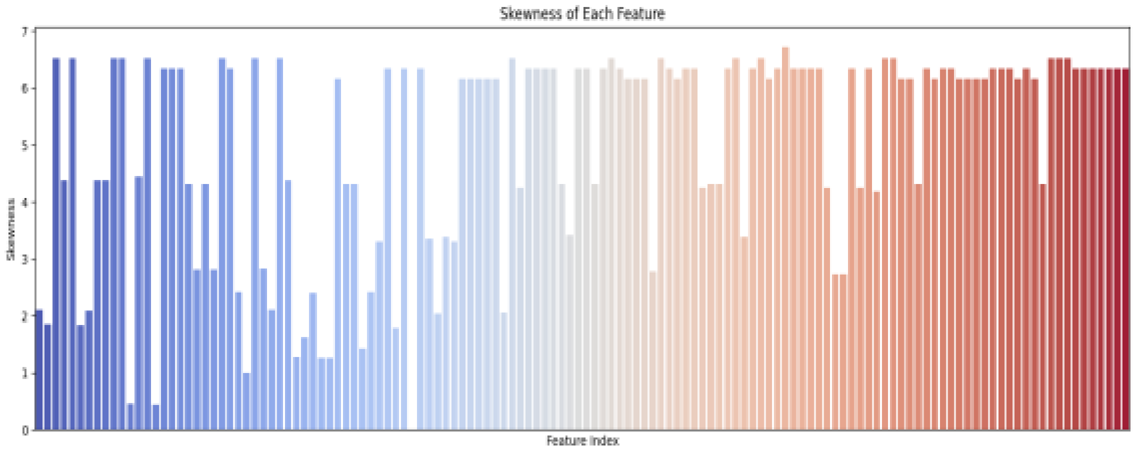


Fig -2 Skewness of Each Feature

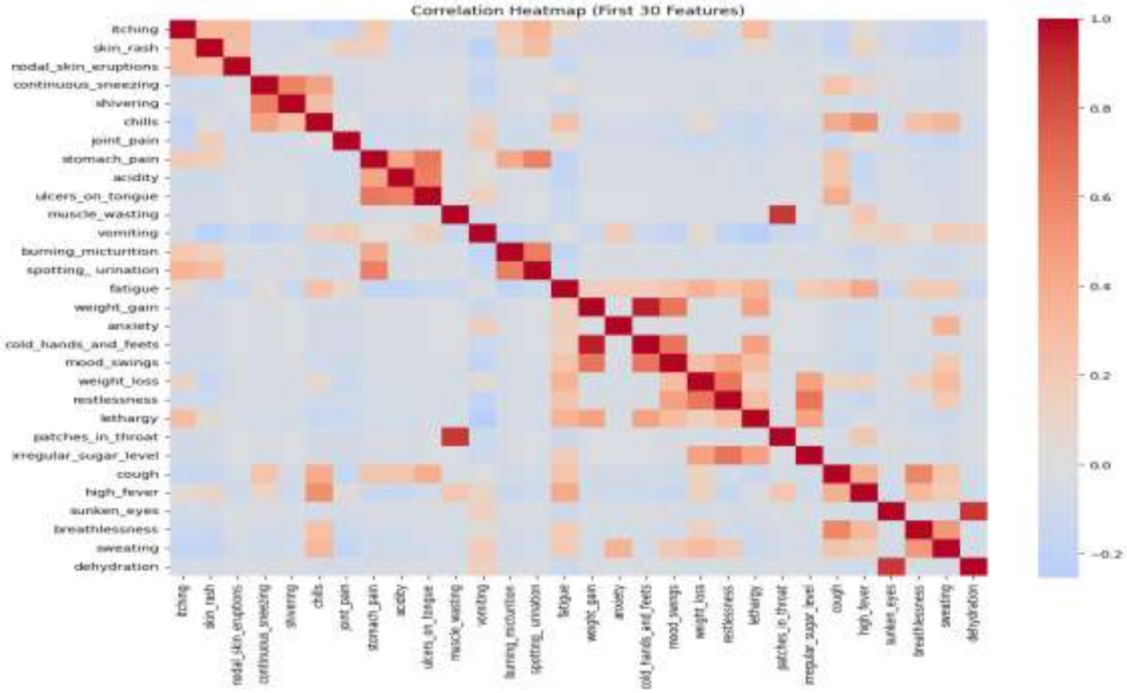


Fig -3 Correlation Heatmap (first 30 Features)

5. Proposed Architecture

Federated Learning is a method in machine learning that allows models to be trained on data distributed across multiple devices or servers, without the need to share the actual raw data. Since privacy is a major concern in this approach, the data exchange mechanism is carefully structured to safeguard sensitive information. Below is a summary of how data is transferred in federated learning with an emphasis on maintaining privacy.

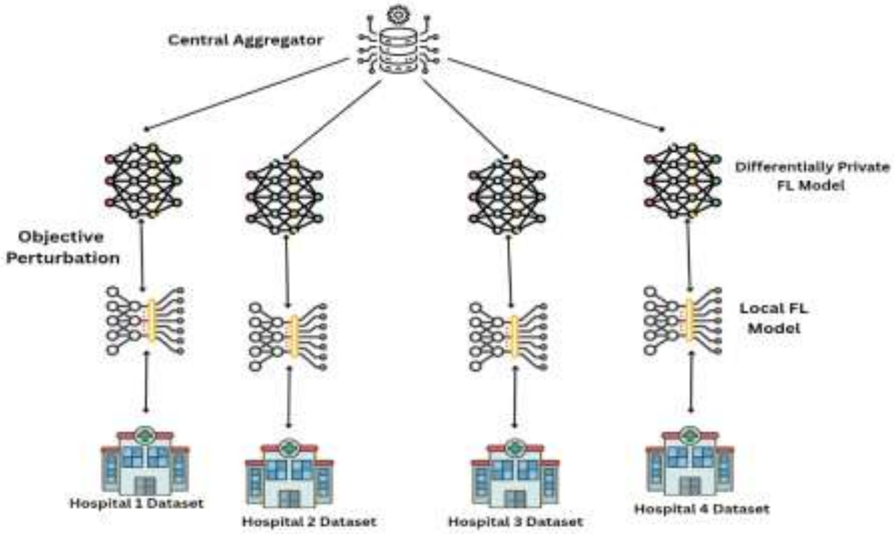
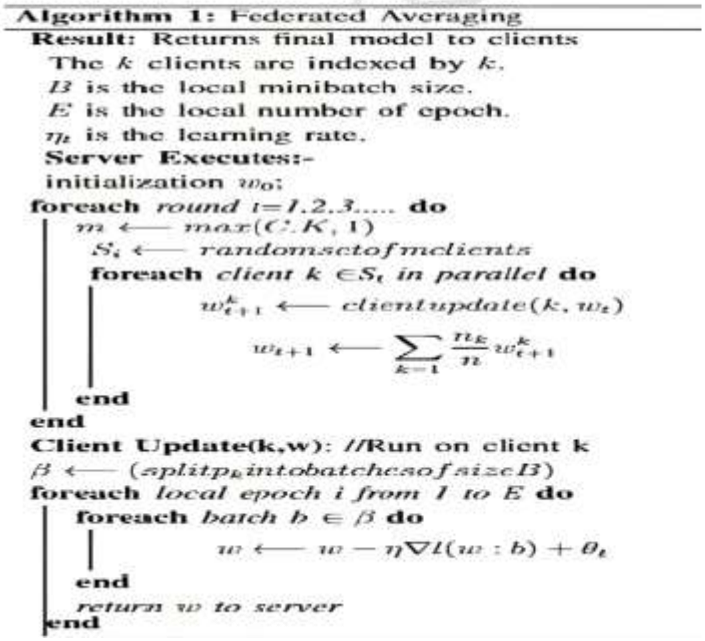


Fig-4 Proposed architecture: Send perturbed model to central aggregator and receive updated central model

In the described model, the process begins with Hospital 1 using its local data to train a model that incorporates differential privacy by adding noise. This privacy-preserving model is then sent to a central aggregator. At the aggregator, the **Federated Averaging (FedAvg)** technique is applied to combine models, resulting in a unified central model. This aggregated model is then sent back to Hospital 1. The same steps are repeated for Hospital 2 and Hospital 3, and the outcomes are evaluated. In each case, the models transmitted from the hospitals include added Gaussian noise to ensure that individual patient records cannot be reverse-engineered or identified.

Algorithm -1 (Federated Averaging) [20]



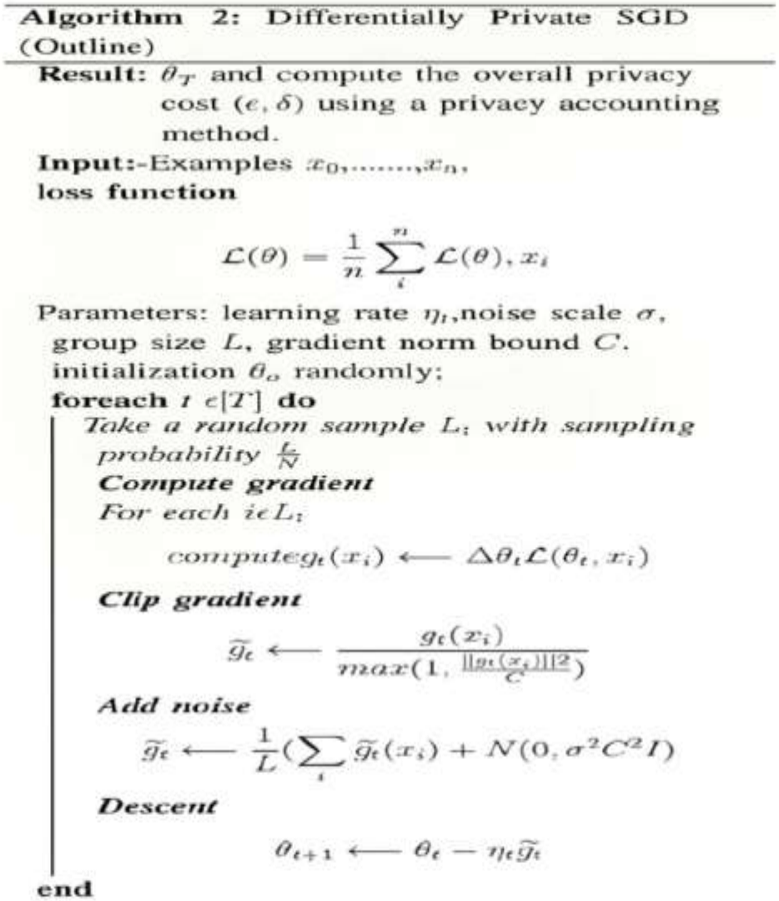
```
Algorithm 1: Federated Averaging
Result: Returns final model to clients
  The $k$ clients are indexed by $k$.
  $B$ is the local minibatch size.
  $E$ is the local number of epoch.
  $\eta_t$ is the learning rate.
  Server Executes:-
  initialization $w_0$;
foreach round $t=1,2,3.....$ do
    $m \longleftarrow max(C.K, 1)$
    $S_t \longleftarrow randomsetofmclients$
    foreach client k $\in S_t$ in parallel do
            $w_{t+1}^k \longleftarrow clientupdate(k, w_t)$
            $w_{t+1} \longleftarrow \sum_{k=1} \frac{n_k}{n} w_{t+1}^k$
    end
end
Client Update(k,w): //Run on client k
$\beta \longleftarrow (splitp_k intobatchesofsizeB)$
foreach local epoch i from 1 to E do
    foreach batch $b \in \beta$ do
            $w \longleftarrow w - \eta \nabla l(w : b) + \theta_t$
    end
    return $w$ to server
end
```

In the federated learning setup, the central aggregator updates the global model by computing a weighted average of the client models after each communication round. To enhance privacy in our proposed approach, we transmit the clients' model parameters to the central aggregator with differential privacy applied. This ensures that the contribution of any individual client remains hidden during the aggregation process [23].

The technique we used to generate differentially private model parameters is illustrated in **Figure 2**, which is based on a variation of Stochastic Gradient Descent (SGD) called **DP-SGD**. This enhanced algorithm modifies the standard mini-batch optimization process by incorporating noise and clipping mechanisms to provide formal differential privacy guarantees [23].

To protect the privacy of each data point within a batch, the algorithm adds Gaussian noise, which helps mask the most dominant gradients. Let's represent **C** as the predefined threshold for the maximum allowable gradient norm. For every data point in the batch, the algorithm first computes its individual gradient. If the gradient's norm is greater than **C**, it is reduced—or "clipped"—so that its norm equals **C**, thereby limiting its influence on the final model update [7].

Algorithm -2 (Differentially Private SGD Algorithm) [20]



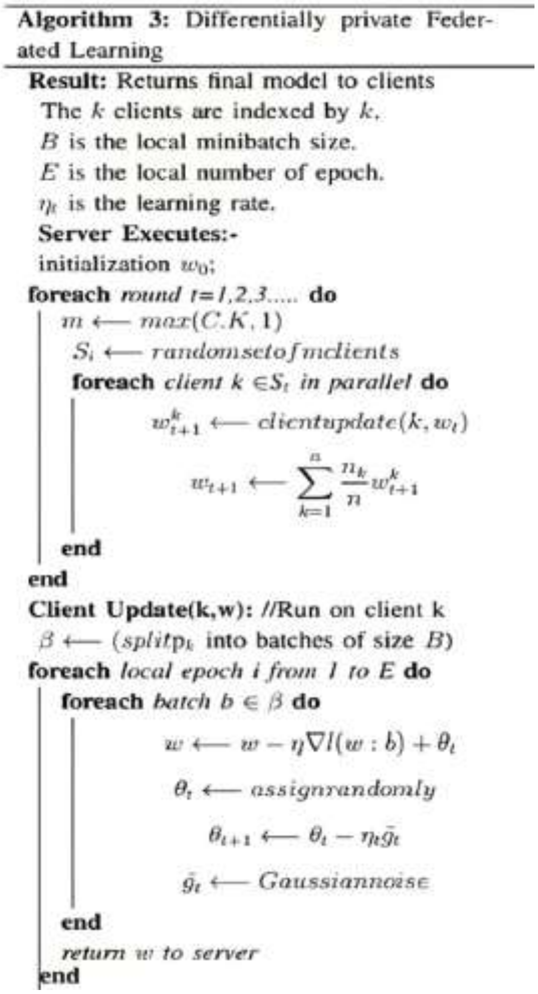
```
Algorithm 2: Differentially Private SGD (Outline)
Result: $\theta_T$ and compute the overall privacy
        cost $(\epsilon, \delta)$ using a privacy accounting
        method.
Input:-Examples $x_0, \ldots\ldots, x_n$,
loss function
        $\mathcal{L}(\theta) = \frac{1}{n}\sum_{i}^{n} \mathcal{L}(\theta), x_i$
Parameters: learning rate $\eta_t$,noise scale $\sigma$,
 group size $L$, gradient norm bound $C$.
 initialization $\theta_o$ randomly;
 foreach $t \in [T]$ do
    Take a random sample $L_t$ with sampling
     probability $\frac{L}{N}$
     Compute gradient
     For each $i \in L_t$
        $compute g_t(x_i) \longleftarrow \Delta\theta_t \mathcal{L}(\theta_t, x_i)$
     Clip gradient
        $\tilde{g}_t \longleftarrow \frac{g_t(x_i)}{max(1, \frac{\|g_t(x_i)\|2}{C})}$
     Add noise
        $\tilde{g}_t \longleftarrow \frac{1}{L}(\sum_i \tilde{g}_t(x_i) + N(0, \sigma^2 C^2 I)$
     Descent
        $\theta_{t+1} \longleftarrow \theta_t - \eta_t \tilde{g}_t$
end
```

Objective perturbation techniques are applied to the locally trained model before it is sent to the central aggregator. The main goal of each client's model is to learn parameters that accurately map input data to outputs by minimizing an associated loss function. To achieve this, **Stochastic Gradient Descent (SGD)** is used to iteratively update the model parameters toward optimal values. To safeguard data privacy, we employ a modified version of **Differentially Private Stochastic Gradient Descent (DP-SGD)**. This method uses a specific update rule, where **C** is the **clipping parameter**, which sets the maximum allowable L2 norm for each gradient. A function is used to adjust (or "clip") any gradient vector whose norm exceeds **C**, ensuring it remains within the limit. Additionally, a **noise multiplier** is applied, which determines how much Gaussian noise is added to each clipped gradient. This multiplier is based on the ratio between the clipping parameter and the standard deviation of the added noise, providing a formal guarantee of differential privacy [24].

$$[x]c=x/(1,||x||2/c)$$

We have adopted a simple strategy where all clients are independent and identically distributed (IID), each operating with its own local model. Before sending their gradient updates to the central aggregator, Gaussian noise is added to each client's gradients to ensure privacy. The central aggregator then applies the **FedAvg** algorithm to aggregate these noisy gradients and build the global model. This final model is subsequently shared with all the IID clients[24].

Algorithm-3 (Differentially Private Federated Learning Algo) [20]



```
Algorithm 3: Differentially private Federated Learning
Result: Returns final model to clients
 The $k$ clients are indexed by $k$.
 $B$ is the local minibatch size.
 $E$ is the local number of epoch.
 $\eta_t$ is the learning rate.
 Server Executes:-
 initialization $w_0$;
foreach round t=1,2,3..... do
    $m \longleftarrow max(C.K, 1)$
    $S_t \longleftarrow randomsetofmclients$
    foreach client $k \in S_t$ in parallel do
        $w^k_{t+1} \longleftarrow clientupdate(k, w_t)$
        $w_{t+1} \longleftarrow \sum_{k=1}^{n} \frac{n_k}{n} w^k_{t+1}$
    end
end
Client Update(k,w): //Run on client k
 $\beta \longleftarrow$ (split$p_k$ into batches of size $B$)
foreach local epoch i from 1 to E do
    foreach batch $b \in \beta$ do
        $w \longleftarrow w - \eta\nabla l(w : b) + \theta_t$
        $\theta_t \longleftarrow assignrandomly$
        $\theta_{t+1} \longleftarrow \theta_t - \eta_t\bar{g}_t$
        $\bar{g}_t \longleftarrow Gaussiannoise$
    end
    return $w$ to server
end
```

To build a privacy-preserving model, we applied an **objective perturbation** method, which adds noise directly to the objective (loss) function before performing optimization over the classifier space. At each client node, **Differentially Private Stochastic Gradient Descent (DP-SGD)** is used to compute gradient updates across mini-batches. During this process, the gradients are first clipped based on a predefined threshold and then noise is added. Specifically, the algorithm takes two inputs: **C**, the clipping threshold, and **σ**, the noise multiplier. It ensures that the L2 norm of each gradient does not exceed **C**, and then adds Gaussian noise with a standard deviation of **σC** to the gradient, providing differential privacy. For model aggregation, the system uses the **Federated Averaging (FedAvg)** algorithm, which computes a weighted average of the differentially private model parameters received from the clients. This aggregated model is then redistributed to all participating nodes to continue the training process [25].

6. Results And Discussion

Figure 5 illustrates the accuracy progression of a machine learning model trained using data from three hospitals—**Hospital-1**, **Hospital-2**, and **Hospital-3**—across 10 training epochs. The **y-axis** represents the accuracy score (ranging from 0 to 0.9), while the **x-axis** represents the number of epochs.

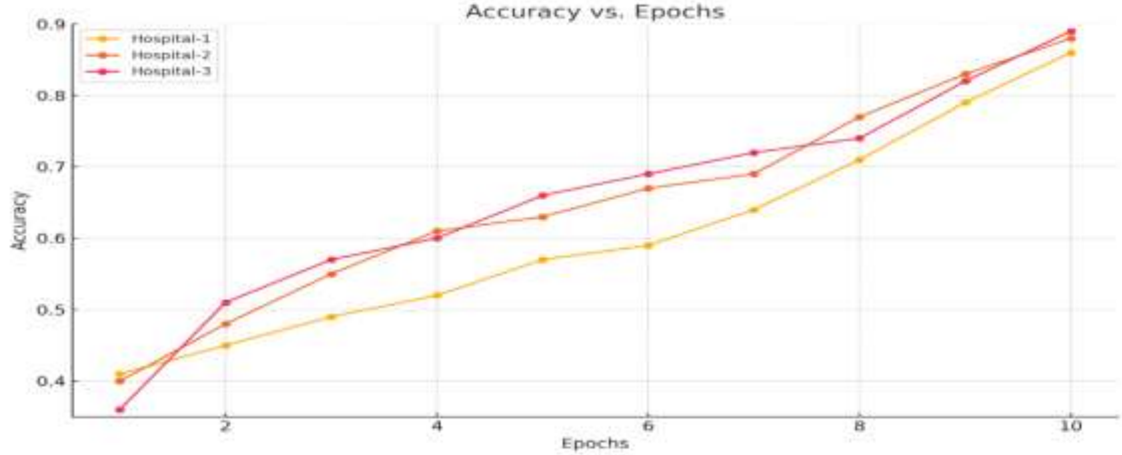


Fig-5 Accuracy of model on individual hospital data

Hospital-1 starts with the lowest accuracy (approximately 0.4) but shows consistent improvement, reaching about 0.8 accuracy by epoch 10.

Hospital-2 begins with a slightly higher starting accuracy than Hospital-1 and demonstrates steady growth, achieving close to 0.85 accuracy at the 10th epoch.

Hospital-3 starts with the highest initial accuracy (around 0.35) and shows the quickest improvement in the early epochs. By the end of the 10th epoch, its accuracy matches that of Hospital-2, at roughly 0.85.

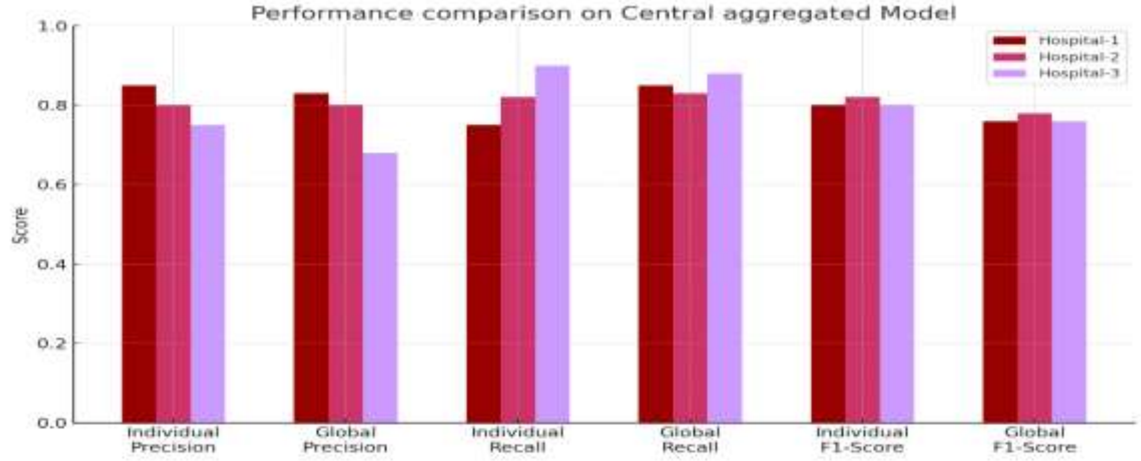


Fig-6 Performance comparison on Central aggregated Model

Figure 5 presents a comparison of the performance of three hospitals—Hospital-1, Hospital-2, and Hospital-3 using two types of models: Individual and Global, evaluated across three key metrics: Precision, Recall, and F1-Score.

Metric Descriptions:

Precision: Indicates how many of the model's positive predictions are actually correct. A higher precision means fewer false positives.

Recall: Reflects the model’s ability to detect all actual positive cases. A higher recall means fewer false negatives.

F1-Score: Combines precision and recall into a single value, offering a balanced view of the model's accuracy in identifying relevant instances.

Comparison of Models:

The Individual model refers to the performance results of each hospital's locally trained model.

The Global model represents a centralized model that aggregates data or insights from all hospitals, aiming to generalize across institutions.

Hospital-1

- Precision: The individual model slightly outperforms the global model in terms of precision.
- Recall: Both models exhibit nearly identical recall values.
- F1-Score: Performance is similar for both models, though the global model has a slight edge.

Hospital-2

- Precision: The global model achieves higher precision than the individual model.
- Recall: The global model shows a modest improvement in recall over the individual model.
- F1-Score: Overall, the global model delivers better performance.
-

Hospital-3

- Precision: The global model demonstrates a clear advantage in precision over the individual model.
- Recall: The recall of the individual model is lower than that of the global model.
- F1-Score: The global model outperforms the individual model across the board.

Overall Observation

Overall, the **global model** demonstrates either enhanced or similar performance across all evaluation metrics and hospitals.

However, the degree of improvement varies by hospital, with **Hospital-3 experiencing the most significant performance boost** from the global model.

While each hospital's **individual model** performs well, none consistently surpasses the performance of the global model.

7. CONCLUSION

In this study, we investigated the application of **differential privacy**, **federated learning**, and the **federated averaging algorithm** as methods to safeguard the privacy of disease-related data. Our experimental findings highlight the practicality and effectiveness of this combined approach in securing sensitive healthcare information. For future research, these methods can be tested on real-world medical datasets to assess their performance in more realistic environments. By implementing these privacy-preserving techniques, it is possible to conduct meaningful analysis on disease data without compromising individual privacy.

REFERENCES

[1]. van Tilborg, H.C.A., Jajodia,Differential Privacy,pp 338-340,2011.

[2]. Dwork, C., & Roth, A.,Foundations and Trends® in Theoretical Computer Science, 211–407,2014.

[3]. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Advances and Open Problems in Federated Learning,Foundations and Trends® in Machine Learning,1-210,2021.

[4]. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. Communication-Efficient Learning of Deep Networks from Decentralized Data.Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS),PMLR 54:1273–1282,2017.

[5]. Kaggle. Diabetic retinopathy detection, 2015.

[6]. Silva, D., Fernandes, L., Silva, F., Ramos, A., & Gomes, J.Federated Learning in Healthcare: Experimental Study with Deep Learning Models Using CloudLab, In *Proceedings of IEEE International Conference on E-health Networking, Application & Services (Healthcom), 1–6.2022.

[7]. Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F,Journal of Healthcare Informatics Research, Journal of Healthcare Informatics Research,1-19,2021.

[8]. Ng, D., Lee, J. Y., Kim, S., Lee, Y., & Yoon, S. Privacy-Preserving Breast Cancer Prediction with Federated Learning andMulti-InstitutionalData,IEEEAccess,117831–117840.2021.

[9]. Substra Foundation, HealthChain Project, Retrieved July 1,2021,from https://www.substra.ai/en/healthchainproject

[10]. Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. IEEE Signal Processing Magazine, 29(6), 141–142.

[11]. Tian Li and Anit Kumar Sahu and Ameet Talwalkar and Virginia Smith (2019). Federated Learning: Challenges, Methods, and Future Directions. CoRR, abs/1908.07873.

[12]. Priyanka Mary Mammen (2021). Federated Learning: Opportunities and Challenges. CoRR, abs/2101.05428.

[13]. J. Jang-Jaccard and S. Nepal, “A survey of emerging threats in cybersecurity”, Journal of Computer and System Sciences, vol. 80. 2014, pp.974-999.

[14]. S. Vishnu, S. R. J. Ramson, R. Jegan, “Internet of Medical Things (IoMT) - An overview”, Mar 2020.

[15]. M. M. Nair and A. K. Tyagi and R. Goyal, “Medical Cyber Physical Systems and Its Issues”, vol. 165. Jan 2019.

[16]. “Data Protection” [Internet]. TheICE. [cited 2022 Mar 12] Available from: https://www.theice.com/data-protection.

[17]. Federated Learning, Ekkono Solutions AB. 2020 May.

[18]. M. Hao, H. Li, G. Xu, Z. Liu, and Z. Chen, ``Privacy-aware and resourcesaving collaborative learning for healthcare in cloud computing," in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2020, pp. 1_6.

[19]. R. Shao, H. He, H. Liu, and D. Liu, ``Stochastic channel-based federated learning for medical data privacy preserving," 2019, arXiv:1910.11160.

[20]. [Mr. Kanhaiya Jee Jha1, Dr. Gaurav Kumar Ameta2, Dr. Esan P Panchal3, Keyurbhai A. Jani4, Pramod Tripathi5, Dr. Shruti B. Yagnik, Privacy- Enhanced Fungal Infection Detection: Leveraging Differential Privacy and Federated Learning in Healthcare System, Journal of Neonatal Surgery ISSN(Online): 2226-0439 Vol. 14, Issue 2 (2025).

[21]. B. Liu, B. Yan, Y. Zhou, Y. Yang, and Y. Zhang, ``Experiments of federated learning for COVID-19 chest X-ray images," Tech. Rep., 2020.

[22]. K. Ogata, Discrete-Time Control Systems. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.

[23]. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, andV. Smith, ``Federated optimization in heterogeneous networks," 2018, arXiv:1812.06127.

[24]. A. F. Agarap, ``Deep learning using recti_ed linear units (ReLU)," 2018, arXiv:1803.08375.

[25]. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, “Federated learning with non -iid data,” arXiv preprint arXiv:1806.00582, 2018.