Federated Learning Implementation with Privacy Leakage Prevention for Hand-Written Digit Recognition

Original Scientific Paper

N. Indira Priyadarsini
Department of Computer Science Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, AP, India
nethalapriya@gmail.com

Dr G. Raja
Department of Computer Science Engineering,
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India
rajajjcet06@kluniversity.in

Abstract *– Federated learning (FL) has brought significant advantages to applications where collaborative learning should occur at multiple participating devices to enhance user experience in specific tasks. However, FL results in privacy leakage when n-1 clients collude to infer the model of another client. In this paper, we not only implement an FL framework but propose a methodology for preventing privacy leakage while realizing machine learning-based automatic hand-written digit recognition. Our framework supports the FL of deep networks where models trained locally are averaged. Two machine learning models Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) are implemented with FL. We proposed an algorithm, Federated Averaging with Privacy Leakage Prevention (FA-PLP), for model averaging to be done by the server. Our algorithm exploits differential privacy (DP) for realizing model averaging while getting rid of chances of privacy leakage. We evaluated our framework with two distributions of the MNIST dataset. Our empirical results revealed that FA-PLP with the CNN model could achieve the highest accuracy of 95.38%.*

Keywords*:* *Federated Learning, Machine Learning, Deep Learning, Privacy, Collaborative Machine Learning*

Received: December 16, 2023; Received in revised form: March 22, 2024; Accepted: April 1, 2024

1. INTRODUCTION

Federated learning (FL) is a novel phenomenon in which multiple distributed clients are involved in the machine learning process collaboratively while preserving the privacy of locally available training data. Though FL minimizes privacy risk, it still may cause leakage of information about local training data in terms of the model's parameters or weights. Therefore, it is indispensable to overcome this problem by proposing algorithms to realize ML models while preserving privacy. With the emergence of fog computing and edge computing, it is made possible for diversified computing devices can participate in the FL process. For instance, modern smartphones when involved in FL can result in a rich user experience [1]. FL enables ML models to be trained in remote clients while localizing training data. A real-world example for FL is that in the healthcare domain, many hospitals (clients) can collaboratively participate in training a model to leverage prediction accuracy for a given disease diagnosis. FL assumes significance when the clients are not willing to share their training data due to locally prevailing privacy policies.

Many research endeavours are found in the literature on FL. Tao et al. [2] address privacy concerns in Vehicular Edge Computing (VEC) with Federated Learning (FL) in autonomous driving, considering malicious parties. Kang et al. [3] introduced FedGRU, a federated learning-based traffic flow prediction algorithm that maintains privacy while achieving accurate predictions. Zhao et al. [4] proposed a smart home system using federated learning (FL) and a reputation mechanism to help home appliance manufacturers improve their products. Zhang et al. [5] introduced VFL, a privacy-preserving and verifiable federated learning method for big data in industrial IoT, enabling effective verification with constant overhead. Fang et al. [6] introduced an efficient, privacy-preserving federated learning

(HFWP) scheme for cloud computing. It is observed from the literature that FL has significant limitations such as probably insecure communication and privacy leakage. Privacy leakage occurs when n-1 clients collude to infer the model of another client. In this paper, we focus on proposing a framework which addresses privacy concerns in FL. Our contributions to this paper are as follows.

1. We proposed an FL framework along with a methodology for preventing privacy leakage while realizing machine learning-based automatic handwritten digit recognition.
2. We proposed an algorithm known as Federated Averaging with Privacy Leakage Prevention (FA-PLP) for model averaging to be done by the server. It addressed the problem of n-1 clients colluding to infer the model of another client (privacy leakage).
3. We built an application to evaluate our FL framework using machine learning techniques like CNN and MLP, for automatic handwritten digit recognition, on two data distributions.

The remainder of the paper is structured as follows. Section 2 reviews existing FL methods and their limitations. Section 3 presents the proposed FL framework with underlying mechanisms and algorithms. Section 4 presents the results of our experiments with two data distributions. Section 5 concludes our work and provides directions for the future scope of the research.

2. RELATED WORK

This section reviews existing methods on FL. Chunyi et al. [1] proposed a fog computing scheme to enhance federated learning, bolstering IoT data privacy and security against various attacks. Demonstrated efficiency and potential for further improvements. Li et al. [2] address privacy concerns in Vehicular Edge Computing (VEC) with Federated Learning (FL) in autonomous driving, considering malicious parties. FL improves training efficiency and privacy, reducing training loss by 73.7% and enhancing accuracy in simulations under different scenarios. The proposed system significantly reduces bandwidth requirements.

Yi et al. [3] introduced FedGRU, a federated learning-based traffic flow prediction algorithm that maintains privacy while achieving accurate predictions. It outperforms state-of-the-art methods in privacy preservation, demonstrating minimal accuracy loss. In Further the work is to enhance prediction accuracy using a Graph Convolutional Network (GCN).

Yang et al. [4] proposed a smart home system using federated learning (FL) and a reputation mechanism to help home appliance manufacturers improve their products. The system involves two stages: customers train an initial model provided by the manufacturer using mobile phones and edge computing. Differential privacy protects features and ensures privacy. The proposed approach guarantees accuracy and data privacy. Anmin et al. [5] introduced VFL, a privacy-preserving and verifiable federated learning method for big data in industrial IoT, enabling effective verification with constant overhead. Experimental results support its efficiency. Chen et al. [6] introduced an efficient, privacy-preserving federated learning (HFWP) scheme for cloud computing. It employs lightweight encryption and optimization strategies. The approach is secure, improves efficiency, and is suitable for cloud and fog computing applications, offering possibilities for further research, including combining SMC with DP and exploring alternative SMC techniques like Pallier. The private leakage prevention approach in FL in the proposed methodology in this paper is different from [6] in both client-side and server-side phenomena besides in the usage of differential privacy.

Zhao et al. [7] proposed a privacy-preserving federated learning approach for industrial big data. It minimizes parameter sharing, uses differential privacy with a Gaussian mechanism, a proxy server for anonymity, and a self-stop mechanism to enhance privacy while maintaining accuracy and performance and It is also related to the previous article.

Yu et al. [8] proposed a privacy-preserving federated learning scheme that ensures both privacy and integrity through a Trusted Execution Environment (TEE). This scheme addresses causative attacks, making collaborative deep learning more secure and practical. It aims to bring the benefits of deep learning to domains with privacy and availability concerns.

Elgabli et al. [9] proposed an analog-based federated learning framework, that addresses wireless channel challenges to improve privacy, bandwidth efficiency, and scalability. It uses analogue transmissions, preserving data privacy and demonstrating effectiveness under various conditions. Major contributions include theoretical advancements and algorithmic innovations. Yang et al. [10] introduced an asynchronous federated learning (AFL) framework for multi-UAV networks, allowing local model training without transmitting raw data. It employs device selection and an A3C-based algorithm to improve learning accuracy and speed. Simulations confirm its superior performance. Yunlong et al. [11] presented a blockchain-based secure data-sharing system for Industrial IoT, integrating federated learning into permissioned blockchain for data privacy and efficiency. Numerical results validate its effectiveness. Future work should explore further security threats, enhance data model utility, and address resource constraints in IIoT data sharing. Xiaoxiao et al. [12] introduced a privacy-preserving federated learning framework for multi-site fMRI analysis, overcoming privacy concerns and enhancing neuroimage analysis. It offers potential benefits in other medical data analysis fields. The approach allows data from various institutions to be utilized while safeguarding privacy, and fostering collaboration in medical research.

Xiaofeng et al. [13] explored real-time data sharing for smart cities must ensure privacy. An adaptive pseudonymization framework enhances privacy robustness in real-time information brokering, with early positive results. Future work includes comprehensive validation and consideration of potential multi-dimensional correlation attacks. The approach could be applied to various information sources beyond energy data. Islam [14] focused on enhancing Federated Learning (FL) for Electronic Health Records (EHRs) by ensuring privacy through techniques such as data generalization, feature selection, and noise minimization. A distributed framework is proposed where local models make predictions based on local features, with added privacy protection using differential privacy. Weighted feature functions ensure a balanced trade-off between privacy and utility. No raw data, features, or model parameters are shared. The method aims to maintain data localization and can be applied to healthcare data, with the potential for future comparisons and improvements.

Chamikara et al. [15] introduce a distributed perturbation algorithm called DISTPAB, addressing privacy concerns in distributed machine learning for geographically dispersed data, like healthcare and banking. DISTPAB shows minimal utility degradation and serves as a promising privacy preservation method for distributed machine learning. Future work will explore further efficiency improvements, particularly in the context of vertical federated learning with varying feature spaces.

Zhang et al. [16] discussed federated learning for privacy-preserving medical models in IoT-based healthcare. It uses cryptographic techniques and data quality weighting. The proposed scheme maintains privacy, and the experiments indicate promising accuracy for lesion cell type detection. In future, it includes optimizing for heterogeneous environments and addressing malicious server issues.

Jiang et al. [17] introduced PFLM, a privacy-preserving federated learning scheme with membership proof, addressing the dropout constraint while ensuring security and verifiability. Security analysis and experiments confirm its efficiency. Yuanhang et al. [18] found that a blockchain-based federated learning system ensures secure and privacy-preserving traffic flow prediction by decentralizing model updates and applying differential privacy. Yin et al. [19] proposed a novel hybrid privacy-preserving federated learning approach that uses advanced encryption, noise addition, and sparse differential gradients to enhance security and efficiency. Yunlong et al. [20] describe an intelligent, secure architecture and privacy-preserving federated learning in VCPS to combat data leakage effectively, ensuring accuracy and security. Ma et al. [21] A privacy-preserving Byzantine-robust federated learning scheme (PBFL) enhances robustness and privacy by using encryption and zero-knowledge proof, providing higher privacy protection. Xiaoyuan et al. [22] introduced an Adaptive Privacy-preserving Federated Learning framework with differential privacy. It uses relevance propagation and adjustment technology to optimize the trade-off between accuracy and privacy, demonstrated through formal analysis and experiments.

Shixiang et al. [23] presented CI-PPFL, a class-imbalance privacy-preserving federated learning framework for decentralized wind turbine fault diagnosis. Experiments on real-world data show its superiority and privacy preservation. Future work includes extending it to heterogeneous label subspaces and integrating vibration data and SCADA data for broader applications. Wei et al. [24] observed that UDP algorithm adds artificial noise to shared models in Federated Learning, ensuring user-level differential privacy. CRD method enhances learning efficiency and model quality for specified privacy levels. Future work aims to refine privacy budget allocation.

Ali et al. [25] explored privacy concerns in IoMT by introducing federated learning (FL) as a solution. It surveys privacy issues in IoMT, discusses existing privacy techniques, and emphasizes FL's collaborative, privacy-preserving nature. The survey further explores FL's advanced architectures with DRL, DNN, and GANs. Finally, it suggests real-time applications and future research directions for improving privacy in smart healthcare systems.

Kong et al. [26] focused on privacy-preserving, flexible model aggregation in federated learning-based automotive navigation called FedLoc. Extensive analysis demonstrates its privacy and security properties, along with improved computational efficiency during participant changes. Future work includes real-world testing and performance assessment. Han et al. [27] proposed a verifiable federated learning scheme is for deep neural networks. It addresses privacy, trust, and accuracy concerns using key exchange, double masking, and tag aggregation. Security and efficiency analyses confirm its effectiveness. Fang et al. [28] introduced PCFL, a privacy-preserving, communication-efficient federated learning approach for IoT. PCFL excels in communication efficiency and model accuracy. Future work targets multi-task learning and advanced cryptographic protocols for IoT security. Tian et al. [29] explored federated learning's unique attributes and challenges, highlighting its distinct nature compared to traditional machine learning. It provides an overview of current approaches and identifies areas for future interdisciplinary research. Cheng et al. [30] introduced SecureBoost, a privacy-preserving tree-boosting system in the context of federated learning, offering accuracy comparable to non-private methods. Information leakage is analysed, and solutions are suggested.

Huafei et al. [31] studied privacy-preserving weighted federated learning within a secret-sharing framework. It introduces weighted federated learning (wFL) and presents its implementation using random splitting and ElGamal encryption. The proposed solution is secure against honest-but-curious adversaries.

Wang et al. [32] introduced VANE, a secure and non-interactive federated learning scheme for regression training with gradient descent. VANE facilitates training global regression models while preserving data privacy. It features a secure data aggregation algorithm and improved training efficiency. Security analysis and experiments demonstrate its effectiveness. Li et al. [33] reviewed the evolution of Federated Learning (FL) in industrial engineering and computer science. It identifies research fronts, summarizes applications, and outlines FL's development prospects. This comprehensive analysis aims to guide future applications and address remaining challenges in FL. Lakhan et al. [34] discussed privacy and fraud issues in machine-learning-based Internet of Medical Things (IoMT) systems. It introduces the FL-BETS framework, focusing on healthcare applications with energy and delay constraints. FL-BETS outperforms existing models. Future work aims to address mobility fraud and extend security measures. Jie et al. [35] stated that the proliferation of healthcare data offers significant potential for improving care, but privacy challenges and data fragmentation persist. This survey reviews federated learning technologies, including their application in healthcare, addressing statistical, system, and privacy challenges. Challenges such as data quality and standardization in healthcare data are also discussed.

Liu et al. [36] observed that edge computing is a technology to extends cloud services to the network edge, raises privacy concerns with user data transmission. P2FEC integrates federated learning and edge computing to preserve privacy and build deep learning models without central data storage, outperforming standard edge computing in privacy protection. Future work is to enhance protection against privacy-sensitive data leakage. Zengpeng et al. [37] introduced a triple-band cylindrical dielectric resonator antenna (CDRA) with HEM11, TM01, and HEM12 modes excited simultaneously using a composite feeding structure. Diverse radiation patterns make it suitable for various wireless applications, including WiMAX and vehicular use. Wang et al. [38] discussed the privacy issues in federated learning, particularly in ternary federated learning (TernGrad). This innovative approach improves communication efficiency and accuracy, representing the first research combining ternary federated learning with privacy-preserving technologies. Future work includes enhancing efficiency and security. Wei et al. [39] proposed NbAFL, a differential privacy-based approach in federated learning to enhance privacy, involving noise, trade-offs, simulations, and future considerations. Aledhari et al. [40] provided a comprehensive study of Federated Learning (FL), highlighting its importance, enabling technologies, and challenges. It explores real-life applications and suggests directions for the future. FL holds the potential to improve data handling and privacy, but challenges such as fault tolerance, performance, and fairness need addressing in its implementation. From the review of literature issues like privacy and security in communications were still found possible. In this paper, we focus on proposing a framework which addresses privacy concerns.

3. PROPOSED FRAMEWORK

This section presents the system model, problem definition, our methodology for federated learning implementation with privacy leakage prevention for hand-written digit recognition and the proposed algorithm.

3.1. SYSTEM MODEL AND PROBLEM STATEMENT

Let us consider a distributed environment where multiple mobile devices participate in language modelling tasks to recognize hand-written digits. All participating mobile devices train an ML model in a collaborative fashion. Each device trains a model ΔWi locally instead of sending its training data to a remote server. Therefore, each mobile device is known as a client which needs to communicate with the server to send local model to it. The server is responsible for computing a global model send it back to each client. The training process is repeated until it reaches a stopping condition or convergence. The system model with the FL approach is illustrated in Fig. 1.

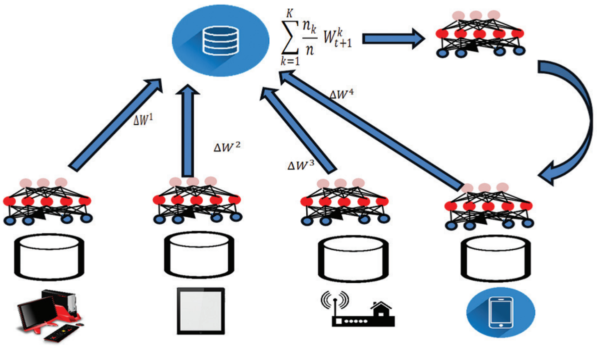

Fig. 1. Illustrates our system model for federated learning

An important advantage of FL is that it is able to decouple model training process and gets rid of direct access to training data. However, the server is essential to coordinate the training process. Therefore, it is essential to have trust in the server or assume it. Nevertheless, there is privacy achieved due to the non-sharing of locally available training data. Thus FL has the potential to minimize security and privacy risks as the attack surface is reduced to the device instead of the attack surface encompassing to entire environment, probably, including the cloud. FL is found ideal for solving many kinds of problems that share common qualities such as distributed availability of data in multiple devices, data is privacy-sensitive and supervised learning where labels can be interactively inferred. However, FL has significant limitations such as probably insecure communication and privacy leakage.

Privacy leakage occurs when n-1 clients collude to infer model of another client. The former (security problem) can be overcome by implementing a secure multiparty communication (MPC) system while the latter (privacy leakage) can be implemented with differential privacy (DP) at each client. In this paper, we focused on FL with privacy-preserving model training through DP implementation.

3.2. OUR METHODOLOGY

The proposed methodology for FL with privacy leakage prevention is based on the system model illustrated in Fig. 1. We considered the problem of privacy leakage which occurs when n-1 clients collude to infer the model of another client. Federated learning, due to its modus operandi, has specific privacy advantages. However, privacy leakage occurs when n-1 clients collude to infer the model of another client. In the FL task, there is a minimal update required to improve model. Privacy depends on the content that needs to be updated in the learning process. Nevertheless, the updates are generally minimal and the source of the update is not required by the aggregation process. Still there is the probability of n-1 clients colluding to cause privacy leakage. Our implementation overcomes this issue as it takes care of privacy-preserving model training. We combine FL with differential privacy to ensure the prevention of privacy leakage in FL. An asynchronous scheme is considered for updates while proceeding with federated communication. A number of clients involved in FL I fixed and their local dataset is also fixed. When each round starts, a fraction of clients are chosen randomly and a global state is obtained from the server. The notion of selecting a fraction of clients is to improve efficiency. Clients perform computation locally on the locally available dataset depending on the global state provided by the server. The result of local computation is sent to the server. Afterwards, the server uses the updates to modify the global state and this procedure is done repeatedly. We considered a finite-sum-based objective for FL as expressed in Eq. 1.

$$\min_{\mathrm{w}\in\mathrm{R}^{\mathrm{d}}} \mathrm{f(w)} \qquad \text{where} \qquad \mathrm{f(w)} \overset{\mathrm{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f_i(w) \tag{1}$$

For given ML problem, f_i $(w)=l(x_i,y_i,w)$ is considered where w denotes model parameters and (x_i,y_i) is the given example on which loss is computed. Assuming that there are k number of clients and data is partitioned accordingly consisting of indexes P_k associated to data in client k and $n_k=|P_k|$ where P_k is the partition. Then the objective can be modified as expressed in Eq. 2.

$$\mathrm{f(w)} = \sum_{\mathrm{k}=1}^{\mathrm{K}} \frac{\mathrm{n}_k}{\mathrm{n}} \mathrm{F}_k(w) \text{where } \mathrm{F}_k(w) = \frac{1}{\mathrm{n}_k}\sum_{i\in\mathrm{P}_k} \mathrm{f}_i(w) \tag{2}$$

P_k is the partition associated with training examples for different clients distributed randomly, it forms the expression $E(P_k)$ $[F_k$ $(w)]=f(w)$. It was observed empirically that in FL communication costs are more than computational costs, unlike the data centre-based approach. In our implementation, each client is involved in less number of updates in FL.

To ensure the prevention of the possibility of privacy leakage in FL, we used differential privacy (DP) which helps in adding noise so as to address privacy attacks. DP is the mathematical model to ensure the privacy of data being exchanged among participants in FL. The DP in its simplest form can be expressed as in Eq. 3.

$$\Pr\,[M(D_1) \in S] \leq e^{\epsilon} Pr[M(D_2) \in S] + \delta \tag{3}$$

When the DP mechanism satisfies expression, it can be used to add noise to the data so as to ensure privacy-preserving communication among clients and servers in FL. It is known as ϵ-differential privacy as discussed in [6]. We consider the Laplacian mechanism that has the potential to preserve ϵ-differential privacy. Considering random noise X, concerning Laplacian distribution, the PFF is expressed in Eq. 4.

$$\mathrm{f(x)}=\frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \tag{4}$$

where λ denotes scale parameter, X is the random noise and the scale value is expressed as in Eq. 5.

$$\lambda = \Delta / \epsilon \tag{5}$$

We also support a distributed approach in adding noise. In this approach, each client adds its portion of noise. Since DP is compatible with the Laplace mechanism, it is possible to generate a Laplace random variable as expressed in Eq. 6.

$$\mathrm{L}(\mu, \lambda)=\frac{\mu}{n} + \sum_{p=1}^{n} \gamma_p - \gamma_p', \tag{6}$$

where γ_p and $\gamma_p{}'$ are random variables as per Gamma distribution, μ and λ denote mean and scale parameters respectively in the Laplace mechanism. Now this leads to the expression in Eq. 7.

$$\frac{(1/s)^{1/n}}{\Gamma(1/n)} x^{\frac{1}{n}-1} e^{-x/s} \tag{7}$$

where s and $1/n$ denote scale and shape parameters respectively. A technique expressed in Eq. 8 is used for each client adding γ_p-$\gamma_p{}'$ in the proposed algorithm which makes use of distributed privacy.

Algorithm 1: Federated Averaging with Privacy Leakage Prevention

```
  Server Side:
  Server initializes w_0
  For each round r in R
        For each client k in K
  $w^k_{r+1}$←ClientSideProcess($k$, $w_r$)
        End For
                    $w_{r+1} \leftarrow \frac{1}{n}\sum_{i=1}^{n} w_{r+1}^{i}$
End For
ClientSideProcess (k, w):
For each local update $u$ in $U$
$w$←$w - \eta\nabla g(w)$
End For
Return $w + \gamma - \gamma'$
```

As presented in Algorithm 1, there are a number founds in which communication takes place between servers and clients as part of FL. In the process, there is server-side functionality and also client-side functionality. In the local updates about weights, noise is added by each client leading to a distributed approach to noise addition. This has the potential to prevent n-1 clients from colluding to infer models of another client. Thus the proposed algorithm helps in preventing privacy leakage. Each client compares its weight with that of the previous round where the server sends global weights to the client. As each client is contributing to the noise addition, it has a more efficient privacy-preserving mechanism in FL. Moreover, on local convergence, each client can come out of the FL system. Each time a client receives federated weight from the server, it can subtract the DP noise it has contributed for those federated weights. With this modus operandi, the proposed FL achieved privacy by defeating any privacy attacks besides supporting the inherent privacy involved in FL.

4. EXPERIMENTAL RESULTS

We made experiments with our implemented prototype for realizing FL. The MNIST dataset used for the empirical study is collected from [41]. The dataset is partitioned as the number of clients involved in the FL. Two approaches are followed to partition data over clients. The first approach simply shuffles the dataset D and distributes it among clients. We call it as D1. The second approach sorts the dataset D based on the digit label, divides it into a number of shards of a given size and each client is provided with a specified number of shards. This is called D2. Experiments are made with both D1 and D2. Models such as CNN and MLP are used for realizing FL.

Table 2. Parameters of CNN along with their values

Parameter	Description	Value
rounds	Number of training rounds	100
C	Client fraction	0.1
K	Number of clients	100
E	Number of training passes on a local dataset for each round	5
batch_size	Batch size	10
LR	Learning rate	0.01

Table 2 shows the parameters used for the CNN model. It uses 100 clients and 100 training rounds with a learning rate of 0.001 and a batch size of 10.

Table 3. Parameters of MLP along with their values

Parameter	Description	Value
rounds	Number of training rounds	100
C	Client fraction	0.1
K	Number of clients	100
E	Number of training passes on a local dataset for each round	5
batch_size	Batch size	10
LR	Learning rate	0.03

Table 3 shows the parameters used for the MLP model. It uses 100 clients and 100 training rounds with a learning rate of 0.03 and batch size 10.

4.1. DATA VISUALIZATION

The dataset collected from [41] is used for experiments. It is related to hand-written digits. It is widely used in machine learning for language modelling and other related applications.

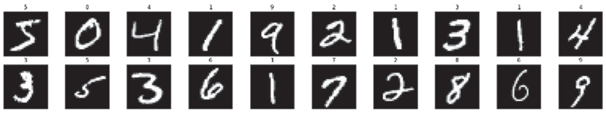

Fig. 2. An excerpt from training data

Fig. 3. An excerpt from test data

Fig. 2 shows an excerpt from training data while Figure 3 presents an excerpt from test data. The dataset is used for hand-written text recognition tasks with FL approach.

4.2. RESULTS OF THE CNN MODEL

Experimental results of FL with CNN model are presented in this section. It provides average loss dynamics and accuracy of the CNN model for two dataset distributions namely D1 and D2.

Table 4. Average loss exhibited by CNN for D1

# Rounds	Average Loss
Round 1	0.818
Round 10	0.04
Round 20	0.026
Round 30	0.02
Round 40	0.018
Round 50	0.014
Round 60	0.013
Round 70	0.013
Round 80	0.009
Round 90	0.008
Round 100	0.006

As presented in Table 4 the average loss exhibited by CNN in FL against different numbers of rounds is provided for D1.

As presented in Fig. 4, the average loss value exhibited by CNN in FL is gradually decreases as the number of rounds is increased. At round 1 the average loss is exhibited as 0.818. The observation at round 10 is reduced to 0.04. When the number of rounds is increased to 50, the average loss value is 0.014. When the number of rounds reaches 100, the average loss observed is the least with 0.006. These observations are recorded when D1 is used for experiments. Less average loss indicates better performance.

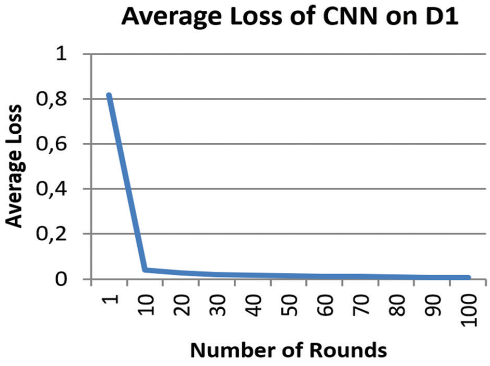

Fig. 4. Average loss of CNN in FL against number of rounds when D1 is used

Table 5. Average loss exhibited by CNN for D2

# Rounds	Average Loss
Round 1	0.097
Round 10	0.021
Round 20	0.017
Round 30	0.008
Round 40	0.012
Round 50	0.007
Round 60	0.006
Round 70	0.006
Round 80	0.006
Round 90	0.006
Round 100	0.004

As presented in Table 5 the average loss exhibited by CNN in FL against different numbers of rounds is provided for D2.

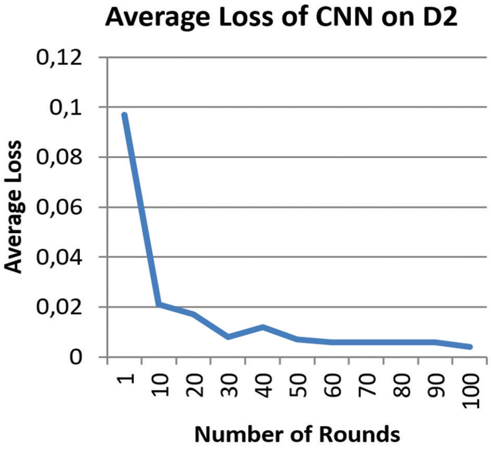

Fig. 5. Average loss of CNN in FL against number of rounds when D2 is used

As presented in Fig. 5, the average loss value exhibited by CNN in FL gradually decreased as the number of rounds is increased. At round 1 the average loss is exhibited as 0.097. The observation at round 10 it is reduced to 0.021. When the number of rounds is increased to 50, the average loss value is 0.007. When the number of rounds reaches 100, the average loss observed is the least with 0.0044. These observations are recorded when D2 is used for experiments.

Table 6. Performance of CNN with FL

Model & Dataset	Accuracy (%)
CNN with D1	95.3856
CNN with D2	94.0512

As presented in Table 6, the performance of the CNN model with the two data distributions is provided in terms of accuracy achieved in hand-written digit recognition.

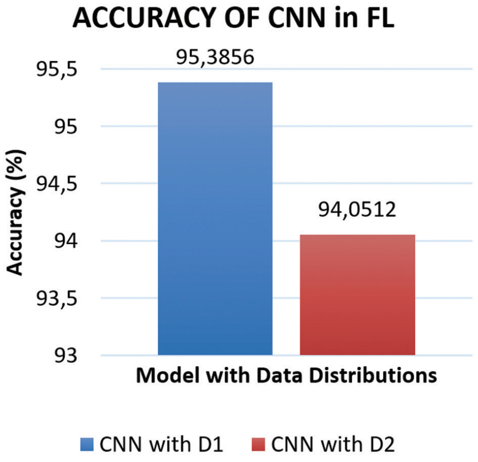

Fig. 6. Accuracy exhibited by CNN with FL when two data distributions are used

As presented in Fig. 6, the accuracy of CNN model in FL with two data distributions is compared. CNN model with D1 achieved better performance with 95.38% accuracy. With D2, the CNN model in FL could achieve 94.05% accuracy.

4.3. RESULTS OF MLP MODEL

Experimental results of FL with MLP model are presented in this section. It provides average loss dynamics and accuracy of the MLP model for two dataset distributions namely D1 and D2.

Table 7. Average loss exhibited by MLP for D1

# Rounds	Average Loss
Round 1	0.607
Round 10	0.059
Round 20	0.032
Round 30	0.026
Round 40	0.027
Round 50	0.017
Round 60	0.018
Round 70	0.013
Round 80	0.01
Round 90	0.013
Round 100	0.008

As presented in Table 7 the average loss exhibited by MLP in FL against different numbers of rounds is provided for D1.

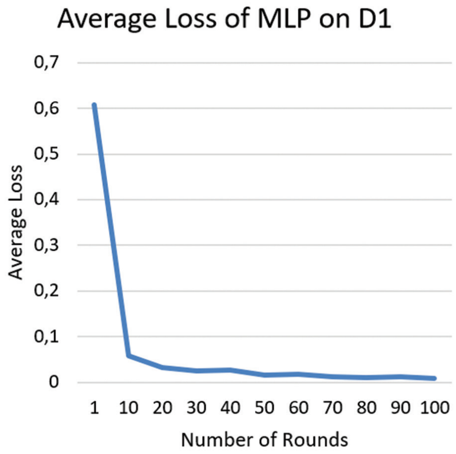

Fig. 7. Average loss of MLP in FL against number of rounds when D1 is used

As presented in Fig. 7, the average loss value exhibited by MLP in FL is gradually decreased as the number of rounds is increased. At round 1 the average loss is exhibited as 0.607. The observation at round 10 it is reduced to 0.059. When the number of rounds is increased to 50, the average loss value is 0.017. When the number of rounds reaches 100, the average loss observed is the least with 0.008. These observations are recorded when D1 is used for experiments. Less average loss indicates better performance.

Table 8. Average loss exhibited by MLP for D2

# Rounds	Average Loss
Round 1	0.125
Round 10	0.022
Round 20	0.013
Round 30	0.011
Round 40	0.014
Round 50	0.005
Round 60	0.008
Round 70	0.004
Round 80	0.012
Round 90	0.006
Round 100	0.008

As presented in Table 8 the average loss exhibited by MLP in FL against different numbers of rounds is provided for D2.

As presented in Fig. 8, the average loss value exhibited by MLP in FL is gradually decreased as the number of rounds is increased. At round 1 the average loss is exhibited as 0.125. The observation at round 10 it is reduced to 0.022. When the number of rounds is increased to 50, the average loss value is 0.005. When the number of rounds reaches 100, the average loss observed is the least with 0.008. These observations are recorded when D2 is used for experiments.

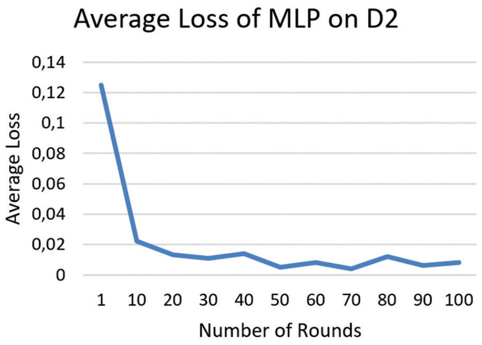

Fig. 8. Average loss of MLP in FL against number of rounds when D2 is used

Table 9. Performance of MLP with FL

Model & Dataset	Accuracy (%)
MLP with D1	93.3216
MLP with D2	90.4032

As presented in Table 9, the performance of the MLP model with the two data distributions is provided in terms of accuracy achieved in hand-written digit recognition.

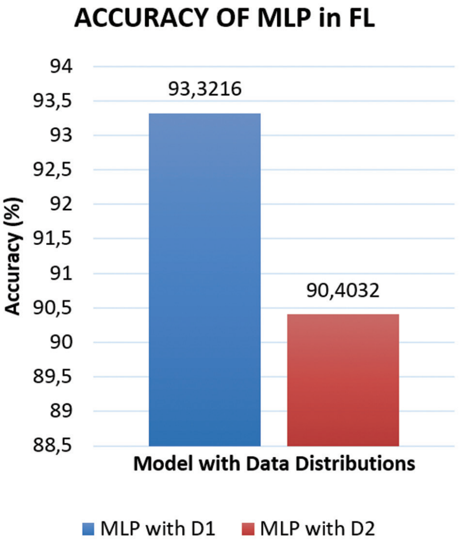

Fig. 9. Accuracy exhibited by MLP with FL when two data distributions are used

As presented in Fig. 9, the accuracy of the CNN model in FL with two data distributions is compared. MLP model with D1 achieved better performance with 93.32% accuracy. With D2, the MLP model in FL could achieve 90.40% accuracy.

4.4. PERFORMANCE COMPARISON

The performance of MLP and CNN models in FL is evaluated in terms of accuracy. The observations are made in this section with two data distributions.

As presented in Table 10, a performance comparison between MLP and CNN in FL is made in terms of accuracy in handwritten digit recognition.

Table 10. Performance comparison between MLP and CNN in FL

Model & Dataset	Accuracy (%)
MLP with D2	90.4032
MLP with D1	93.3216
CNN with D2	94.0512
CNN with D1	95.3856

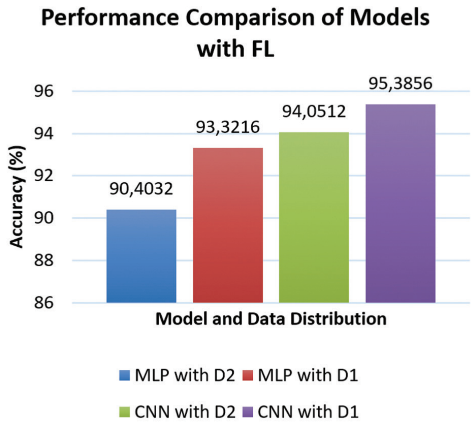

Fig. 10. Accuracy exhibited by MLP and CNN with FL when two data distributions are used

As presented in Fig. 10, the two models such as MLP and CNN are used in FL with two data distributions. The accuracy of MLP with D2 is 90.40%, MLP with D1 93.32%, CNN with D2 94.05% and CNN with D1 95.38%. Highest accuracy achieved by the CNN model with D2 is 95.38%.

Table 11. Performance comparison with state-of-the-art

FL Model	Accuracy (%)
Chen et al. [42]	93.2145
Ng et al. [43]	93.4231
FA-PLP (Proposed)	95.3856

Our results are compared with state-of-the-art methods such as Chen et al. [42] and Ng et al. [43] as presented in Table 11.

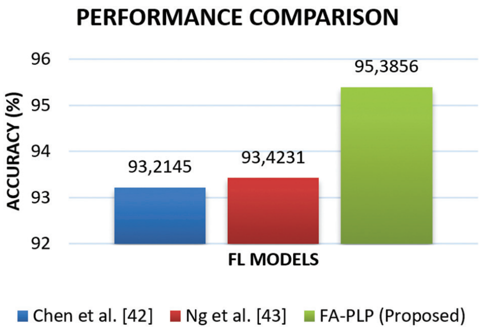

Fig. 11. Performance comparison of FL models

The performance of the proposed model named FA-PLP is compared against existing models. The results revealed that FP-PLP outperforms other models in terms of accuracy with 95.3866%.

5. CONCLUSION AND FUTURE WORK

In this paper, we not only implement an FL framework but propose a methodology for preventing privacy leakage while realizing machine learning-based automatic hand-written digit recognition. Our framework supports the FL of deep networks where models trained locally are averaged. Two models Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) are implemented with FL. We proposed an algorithm, Federated Averaging with Privacy Leakage Prevention (FA-PLP), for model averaging to be done by the server. Our algorithm exploits differential privacy (DP) for realizing model averaging while getting rid of chances of privacy leakage. We evaluated our framework with two distributions of the MNIST dataset. Our empirical results revealed that FA-PLP outperforms existing FL techniques in terms of communication cost, accuracy and privacy leakage prevention. Our framework with the CNN model could achieve the highest accuracy of 95.38%. In future, we intend to improve our framework further by considering the security concerns of FL as well. We also elaborate on different privacy attack scenarios and system behaviours in our future research.

6. REFERENCES

[1] Z. Chunyi, F. Anmin, Y. Shui, Y. Wei, W. Huaqun, Z. Yuqing, "Privacy-Preserving Federated Learning in Fog Computing", IEEE Internet of Things Journal, Vol. 7, No. 11, 2020, pp. 10782-10793.

[2] Y. Li, X. Tao, X. Zhang, J. Liu, J. Xu, "Privacy-Preserved Federated Learning for Autonomous Driving", IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 7, 2022, pp. 8423-8434.

[3] L. Yi, Y.J. Q. James, K. Jiawen, N. Dusit, Z. Shuyu, "Privacy-preserving Traffic Flow Prediction: A Federated Learning Approach", IEEE Internet of Things Journal, Vol. 7, No. 8, 2020, pp. 7751-7763.

[4] Z. Yang, Z. Jun, J. Linshan, T. Rui, N. Dusit, L. Zengxiang, L. Lingjuan, L. Yingbo, "Privacy-Preserving Blockchain-Based Federated Learning for IoT Devices", IEEE Internet of Things Journal, Vol. 8, No. 3, 2020, pp. 1817-1829.

[5] F. Anmin, Z. Xianglong, X. Naixue, G. Yansong, W. Huaqun,Z. Jing, "VFL: A Verifiable Federated Learning with Privacy-Preserving for Big Data in Industrial IoT", IEEE Transactions on Industrial Informatics, Vol. 18, No. 5, 2022, pp. 3316-3326.

[6] F. Chen, G. Yuanbo, W. Na, J. Ankang, "Highly efficient federated learning with strong privacy pres-

ervation in cloud computing", Computers & Security, Vol. 96, 2020.

[7] B. Zhao, K. Fan, K. Yang, Z. Wang, H. Li, Y. Yang, "Anonymous and Privacy-Preserving Federated Learning With Industrial Big Data", IEEE Transactions on Industrial Informatics, Vol. 17, No. 9, 2021, pp. 6314-6323.

[8] C. Yu, L. Fang, L. Tong, X. Tao, L. Zheli, L. Jin, "A Training-integrity Privacy-preserving Federated Learning Scheme with Trusted Execution Environment", Information Sciences, Vol. 522, 2020, pp. 69-79.

[9] A. Elgabli, J. Park, C. B. Issaid, M. Bennis, "Harnessing Wireless Channels for Scalable and Privacy-Preserving Federated Learning", IEEE Transactions on Communications, Vol. 69, No. 8, 2021, pp. 5194-5208.

[10] H. Yang, J. Zhao, Z. Xiong, K. Y. Lam, S. Sun, L. Xiao, "Privacy-Preserving Federated Learning for UAV-Enabled Networks: Learning-Based Joint Scheduling and Resource Management", IEEE Journal on Selected Areas in Communications, Vol. 39, No. 10, 2021, pp. 3144-3159.

[11] L. Yunlong, H. Xiaohong, D. Yueyue, M. Sabita, Z. Yan, "Blockchain and Federated Learning for Privacy-preserved Data Sharing in Industrial IoT", IEEE Transactions on Industrial Informatics, Vol. 16, No. 6, 2019, pp. 4177-4186.

[12] L. Xiaoxiao, G. Yufeng, D. Nicha, S. H. Lawrence, V. Pamela, D. S. James, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results", Medical Image Analysis, Vol. 65, 2020.

[13] L. Xiaofeng, L. Yuying, L. Pietro, H. Pan, "Privacy-Preserving Asynchronous Federated Learning Mechanism for Edge Network Computing", IEEE Access, Vol. 8, 2020, pp. 48970-48981.

[14] T. U. Islam, R. Ghasemi, N. Mohammed, "Privacy-Preserving Federated Learning Model for Healthcare Data", Proceedings of the IEEE 12th Annual Computing and Communication Workshop and Conference, Las Vegas, NV, USA, 26-29 January 2022.

[15] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, S. Camtepe, "Privacy-preserving distributed machine learning with federated learning", Computer Communications, Vol. 171, 2021, pp. 112-125.

[16] L. Zhang, J. Xu, P. Vijayakumar, P. K. Sharma, U. Ghosh, "Homomorphic Encryption-based Privacy-preserving Federated Learning in IoT-enabled Healthcare System", IEEE Transactions on Network Science and Engineering, Vol. 5, No. 2, 2022, pp. 2864-2880.

[17] C. Jiang, C. Xu, Y. Zhang, "PFLM: Privacy-preserving federated learning with membership proof", Information Sciences, Vol. 576, 2021, pp. 288-311.

[18] Q. Yuanhang, H. M. Shamim, N. Jiangtian, L. Xuandi, "Privacy-preserving blockchain-based federated learning for traffic flow prediction", Future Generation Computer Systems, Vol. 117, 2021, pp. 328-337.

[19] L. Yin, J. Feng, H. Xun, Z. Sun, X. Cheng, "A Privacy-Preserving Federated Learning for Multiparty Data Sharing in Social IoTs", IEEE Transactions on Network Science and Engineering, Vol. 8, No. 3, 2021, pp. 2706-2718.

[20] L. Yunlong, H. Xiaohong, D. Yueyue, M. Sabita, Z. Yan, "Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems", IEEE Network, Vol. 34, No. 3, 2020, pp. 50-56.

[21] X. Ma, Y. Zhou, L. Wang, M. Miao, "Privacy-preserving Byzantine-robust federated learning", Computer Standards & Interfaces, Vol. 80, 2022.

[22] L. Xiaoyuan, L. Hongwei, X. Guowen, L. Rongxing, H. Miao, "Adaptive privacy-preserving federated learning", Peer-to-Peer Networking and Applications, Vol. 13, 2020, pp. 2356-2366.

[23] L. Shixiang, Z. Gao, Q. Xu, C. Jiang, A. Zhang, X. Wang, "Class-Imbalance Privacy-Preserving Federated Learning for Decentralized Fault Diagnosis With Biometric Authentication", IEEE Transactions on Industrial Informatics, Vol. 18, No. 12, 2022, pp. 9101-9111.

[24] K. Wei, J. Li, M. Ding, M. Chuan, H. Su, B. Zhang, H. V. Poor, "User-Level Privacy-Preserving Federated Learning: Analysis and Performance Optimization", IEEE Transactions on Mobile Computing, Vol. 21, No. 9, 2021, pp. 3388-3401.

[25] M. Ali, F. Naeem, M. Tariq, G. Kaddoum, "Federated Learning for Privacy Preservation in Smart Healthcare Systems: A Comprehensive Survey", IEEE

Journal of Biomedical and Health Informatics, Vol. 27, No. 2, 2022, pp. 1-10.

[26] Q. Kong, F. Yin, R. Lu, B. Li, X. Wang, S. Cui, P. Zhang, “Privacy-Preserving Aggregation for Federated Learning-Based Navigation in Vehicular Fog”, IEEE Transactions on Industrial Informatics, Vol. 17, No. 12, 2021, pp. 8453-8463.

[27] G. Han, T. Zhang, Y. Zhang, G. Xu, J. Sun, J. Cao, “Verifiable and privacy-preserving federated learning without fully trusted centres", Journal of Ambient Intelligence and Humanized Computing, Vol. 13, 2021, pp. 1431-1441.

[28] C. Fang Y. Guo, Y. Hu, B. Ma, L. Feng, A. Yin, “Privacy-preserving and communication-efficient federated learning in the Internet of Things", Computers & Security, Vol. 103, 2021.

[29] L. Tian, S. A. Kumar, T. Ameet, S. Virginia, “Federated Learning: Challenges, Methods, and Future Directions”, IEEE Signal Processing Magazine, Vol. 37, No. 3, 2020, pp. 50-60.

[30] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, Q. Yang, “SecureBoost: A Lossless Federated Learning Framework”, IEEE Intelligent Systems, Vol. 36, 2021, pp. 87-98.

[31] Z. Huafei, M. Goh, R. Siow, N. Wee-Keong, “Privacy-Preserving Weighted Federated Learning Within the Secret Sharing Framework”, IEEE Access, Vol. 8, 2020, pp. 198275-198284.

[32] F. Wang, H. Zhu, R. Lu, Y. Zheng, H. Li, “A privacy-preserving and non-interactive federated learning scheme for regression training with gradient descent”, Information Sciences, Vol. 552, 2021, pp. 183-200.

[33] L. Li, F. Yuxi, T. Mike, L. Kuo-Yi, “A review of applications in federated learning”, Computers & Industrial Engineering, Vol. 149, 2020.

[34] A. Lakhan, M. A. Mohammed, J. Nedoma, A. Lakhan, R. Martinek, P. Tiwari, A. Vidyarthi, A. Alkhayyat, W. Wang, “Federated-Learning Based Privacy Preservation and Fraud-Enabled Blockchain IoMT System for Healthcare”, IEEE Journal of Biomedical and Health Informatics, Vol. 27, No. 2, 2023, pp. 664-672.

[35] X. Jie, G. S. Benjamin, S. Chang, W. Peter, B. Jiang, W. Fei, “Federated Learning for Healthcare Informatics”, Journal of Healthcare Informatics Research, Vol. 5, 2020, pp. 1-19.

[36] G. Liu, C. Wang, X. Ma, Y. Yang, “Keep Your Data Locally: Federated-Learning-Based Data Privacy Preservation in Edge Computing”, IEEE Network, Vol. 35, No. 2, 2021, pp. 60-66.

[37] L. Zengpeng, S. Vishal, P. M. Saraju, “Preserving Data Privacy via Federated Learning: Challenges and Solutions”, IEEE Consumer Electronics Magazine, Vol. 9, No. 3, 2020, pp. 8-16.

[38] Y. Dong, X. Chen, L. Shen, D. Wang, “EaSTFLy: Efficient and secure ternary federated learning”, Computers & Security, Vol. 94, 2020.

[39] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, H. V. Poor, “Federated Learning with Differential Privacy: Algorithms and Performance Analysis”, IEEE Transactions on Information Forensics and Security, Vol. 15, 2020, pp. 3454-3469.

[40] M. Aledhari, R. Razzak, R. M. Parizi, F. Saeed, “Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications”, IEEE Access, Vol. 8, 2020, pp. 140699-140725.

[41] The MNIST Database. Retrieved from http://yann.lecun.com/exdb/mnist/ (accessed: 2024)

[42] Y. Chen, X. Sun, Y. Jin, “Communication-efficient federated deep learning With layerwise asynchronous model update and temporally weighted aggregation,” IEEE Transactions on Neural Networks and Learning Systems, Vol. 31, No. 10, 2020, pp. 4229-4238.

[43] J. S. Ng et al. “Joint auction-coalition formation framework for communication-efficient federated learning in UAV-enabled Internet of Vehicles,” IEEE Transactions on Intelligent Transportation Systems, Vol. 22, No. 4, pp. 2326-2344.