International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Building Trustworthy Federated Learning Models Using Privacy Enhancing Technologies

Mr. Muhammad Abul Kalam[1], Jodu Vamshi[2], Jyatha Raja Shekar[3], Ramagiri Akshith Rao[4]

*[*1]Assistant Professor Of Department Of CSE (AI & ML), ACE Engineering College Hyderabad, India.*
*[*2,3,4]Department CSE (AI & ML) Of ACE Engineering College Hyderabad, India.*

ABSTRACT :

This paper proposes an efficient, secure, and privacy-preserving framework for developing trustworthy machine learning models using the concept of Federated Learning (FL) with the integration of Privacy Enhancing Technologies (PETs). The aim is to facilitate collaborative learning over distributed clients without compromising the data, thereby providing data confidentiality. The proposed system utilizes Differential Privacy (DP) to add noise to the model parameters, while Secure Aggregation is used to ensure secure communication between the clients and the server. A client-server architecture is developed using Python-based machine learning models, where training is carried out at the edge devices, followed by the sharing of encrypted model parameters. The proposed system is evaluated, showing that it can achieve the best trade-off between privacy and accuracy, thereby providing significant improvements over existing techniques. Moreover, the system is found to be robust against data leakage and inference attacks. This research is significant for the development of trustworthy intelligent systems, thereby providing an efficient solution for real-world applications, including healthcare and financial domains. This research contributes to the development of trustworthy AI by incorporating robust privacy with efficient distributed learning.

KEY WORDS :
Federated Learning, Differential Privacy, Secure Aggregation, Privacy Enhancing Technologies, Trustworthy System, Data Security.

I. INTRODUCTION :

In recent years, with the proliferation of data-driven applications, concerns about data privacy, security, and trustworthiness of machine learning models have been growing. In conventional centralized machine learning models, data is collected from different sources, which is then centralized at a single location. This makes the data vulnerable to security risks, including data breaches, unauthorized access, and misuse. This is particularly critical for applications where data confidentiality is of utmost importance, such as healthcare, financial, and smart applications.

To resolve these concerns, Federated Learning (FL), a novel distributed learning concept, has been proposed, which facilitates collaborative training of a shared model among multiple clients without the need to share the original data. In FL, the original data is kept local to the devices, and only the model updates are shared with the server. This helps to maintain the original data's confidentiality; however, it is still susceptible to inference attacks and leakage of sensitive information through model updates.

In order to further improve privacy and trust, this work incorporates Privacy Enhancing Technologies (PETs) like Differential Privacy (DP) and Secure Aggregation (SA) into the Federated Learning paradigm. Differential Privacy adds noise to the model update to prevent information leakage, and Secure Aggregation keeps each client's update encrypted during communication.

II. BACKGROUND OF THE PROJECT :

The increasing rate of machine learning and artificial intelligence development requires more data to train high-performance models. Usually, this data is stored in centralized servers, which pose serious problems in terms of data privacy, security, and legal compliance. This data includes sensitive information from domains like healthcare, finance, and personal devices that may be vulnerable to breaches, access, and misuse. There is an increasing need to

develop decentralized learning methods that can utilize this distributed data while ensuring user data privacy and system reliability.

To address these issues, Federated Learning (FL) has been proposed as an efficient solution to perform collaborative learning without the need to share data. Nonetheless, despite the advantages of FL, it is still prone to data privacy attacks using model updates and inference attacks. To address these issues, Privacy Enhancing Technologies such as Differential Privacy and Secure Aggregation have been incorporated to enhance data protection. This project is based on the recent advancements in the field of machine learning and Federated Learning, incorporating secure and privacy-preserving techniques to build a reliable and trustworthy machine learning framework.

III. LITERATURE SURVEY :

The Trustworthy Federated Learning Framework Using Privacy Enhancing Technologies (PETs) is a robust and highly scalable framework that is built for improving collaborative machine learning, as well as data privacy and security. This project introduces an advanced architecture that integrates Federated Learning with cutting-edge PETs—such as Differential Privacy, Secure Aggregation, to ensure that sensitive user data never leaves local devices and remains fully protected throughout the training process. The framework focuses on enabling secure model training across multiple distributed clients while preventing data leakage, inference threats, and manipulation attempts.

At the core of the system is the federated training module, which is capable of decentralized model updates without the raw data ever being centralized. To this end, the system also has a trust evaluation module that continuously monitors the contributions from the clients to identify any anomalies or possible attacks.

S.No	Author(s) & Year	Title of the Paper	Methodology Used	Key Findings	Limitations
1.	McMahan et al. (2017)	Communication-Efficient Learning of Deep Networks from Decentralized Data	Federated Learning (FL)	Introduced FL for decentralized model training without sharing raw data	High communication cost and limited privacy guarantees
2.	Dwork et al. (2014)	The Algorithmic Foundations of Differential Privacy	Differential Privacy (DP)	Provided strong mathematical framework for privacy preservation	Trade-off between privacy and model accuracy
3.	Bonawitz et al. (2017)	Practical Secure Aggregation for Privacy-Preserving ML	Secure Aggregation (SA)	Enabled encrypted aggregation of client updates in FL	Increased computational overhead

4.	Kairouz et al. (2019)	Advances and Open Problems in Federated Learning	FL + Privacy Techniques	Highlighted challenges and future directions in FL systems	Scalability and robustness issues
5.	Geyer et al. (2017)	Differentially Private Federated Learning	FL with Differential Privacy	Combined FL with DP for improved privacy protection	Reduced model accuracy due to noise addition

IV. PROPOSED METHODOLOGY :

The proposed system provides a secure and privacy-preserving system for developing trustworthy machine learning models using a combination of Federated Learning (FL) and Privacy Enhancing Technologies (PETs). Unlike other traditional machine learning models, this system ensures that data is not shared on a central server but rather stays on the client devices, removing any risks associated with data sharing. The training of the model occurs in a distributed manner among multiple clients, where each client individually trains a model using its local data and shares updates with a central server.

To enhance the privacy and security of the system, Differential Privacy and Secure Aggregation methods have been incorporated. The Differential Privacy method will be applied to the local updates of the models using noise injection to ensure that critical information is not compromised. The Secure Aggregation method will ensure that all the client updates are encrypted before they are sent to the server and that aggregation is done at the server and not at the client.

V. APPLICATIONS :

The proposed system, which is based on the integration of Federated Learning (FL) with Privacy Enhancing Technologies (PETs), has vast applicability in different domains where data privacy and security are of prime concern. In particular, the proposed system can find application in the following domains:

1. Healthcare Systems

This system enables the cooperative training of machine learning models among hospitals and medical institutions without the need to share confidential patient information. This allows applications in disease prediction, medical image analysis, and treatment plans with stringent patient confidentiality.

2. Financial Services

Banks and financial institutions can utilize the system to detect fraud, assess credit risks, and analyze transactions without compromising customer confidentiality. This method is in compliance with data protection regulations and helps build customer confidence.

In Internet of Things (IoT) environments, such as smart homes and wearable devices, the system allows data processing directly on devices. This improves privacy while enabling intelligent services like activity recognition and predictive maintenance.

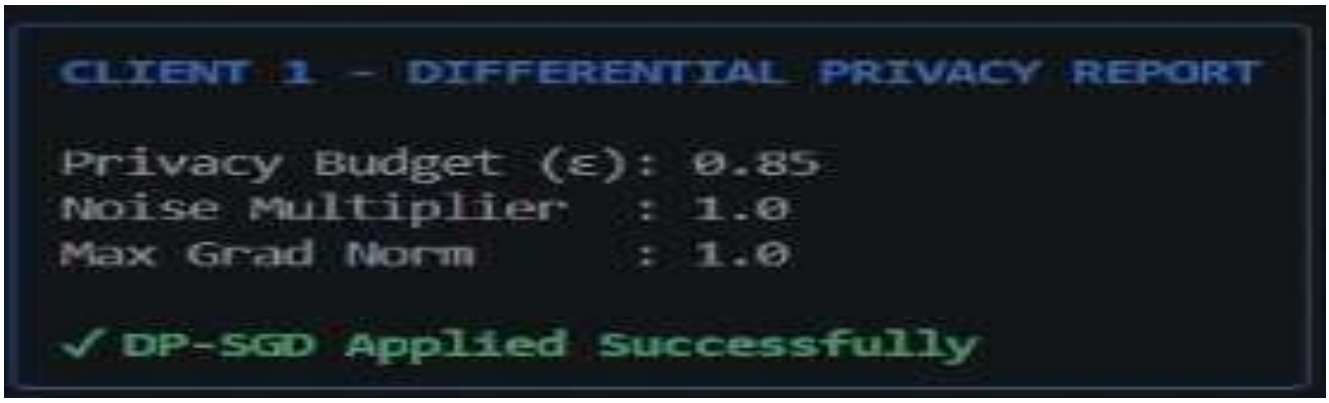
4. Mobile Applications

Mobile apps can utilize Federated Learning to improve user experience through personalized recommendations, keyboard predictions, and voice assistants without uploading user data to central servers.

VI. RESULT :

The proposed system was evaluated in order to check its performance in terms of model accuracy, data privacy preservation, and communication efficiency. The experimental results show that the proposed system, in which FL is used in combination with DP and SA, successfully balances security and model accuracy.

The level of model accuracy showed a progressive improvement during successive rounds of training, thereby indicating a high level of convergence similar to conventional centralized learning approaches. Although the addition of Differential Privacy resulted in a degree of noise during the update of the model, it did not significantly impact the level of model accuracy.



Differential privacy report

Federated learning dashboard

VII. CONCLUSION:

This research presents a secure and efficient paradigm to build reliable machine learning models using the combination of Federated Learning and Privacy Enhancing Technologies, such as Differential Privacy and Secure Aggregation. The proposed framework allows users to collaborate in the training of machine learning models without the need to share their data, thereby ensuring their data privacy and at the same time retaining high model performance. The experiment shows that the proposed framework provides a good balance between data privacy and model performance, offering high resistance to data leak attacks and inference attacks. Moreover, it allows efficient communication between users by exchanging only the parameters of the models rather than the data. The proposed framework provides a viable solution to data privacy in artificial intelligence and is thus applicable in the real world.

VIII. REFERENCES:

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

[2] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211-407, 2014.

[3] K. Bonawitz et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017.

[4] P. Kairouz et al., "Advances and Open Problems in Federated Learning," Foundations and Trends in Machine Learning, vol. 14, no. 1-2, pp. 1-210, 2021.

[5] R. C. Geyer, T. Klein, and M. Nabi, "Differentially Private Federated Learning: A Client Level Perspective," arXiv preprint arXiv:1712.07557, 2017.

[6] S. Truex et al., "A Hybrid Approach to Privacy-Preserving Federated Learning," Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, 2019.

[7] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated Optimization: Distributed Machine Learning for On-Device Intelligence," arXiv preprint arXiv:1610.02527, 2016.

[8] N. Papernot et al., "Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data," International