iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74357

Federated Learning with Differential Privacy: Enhancing Security in Distributed AI Systems

Dr. Diwakar Ramanuj Tripathi[1], Jatin Vitthal waghe[2], Dr. Vrushali Pramod Parkhi[3]
[1]Head, Department of Computer Science, [2]Research Scholar, [3]Officiating Principal, S.S. Maniar College of Computer & Management, Nagpur

Abstract: This paper explores how Federated Learning (FL) systems can be strengthened through the integration of Differential Privacy (DP). While FL allows multiple clients to collaboratively train a shared model without exposing raw data, model updates exchanged during training may still leak sensitive information. To address this, DP is applied using gradient clipping and Gaussian noise addition, thereby reducing the risk of privacy breaches. The study employs the Fed Avg algorithm in simulation experiments with ten clients under three noise levels (σ = 0.0, 0.5, 1.0), evaluating outcomes in terms of accuracy, log loss, and an illustrative Rényi-DP privacy budget (ε). Results highlight the trade-off between privacy and utility: models without noise achieve the highest accuracy but weakest privacy, moderate noise provides balanced performance, and stronger noise enhances privacy at the expense of accuracy. The findings emphasize the importance of tuning parameters such as clipping norm, noise multiplier, communication rounds, and participation rate to balance formal privacy protection with model utility. The study concludes by recommending standardized privacy accounting, randomized client participation, and task-specific parameter tuning as essential practices for securely deploying FL in sensitive domains such as healthcare, finance, and the Internet of Things.

Keywords: Federated Learning, Differential Privacy, Privacy Accounting, Gradient Clipping, Secure Distributed AI.

I. INTRODUCTION

In recent years, federated learning (FL) has drawn a lot of interest as a revolutionary distributed machine learning paradigm. FL permits data to stay with its original owners, in contrast to traditional centralized techniques that need data from several sources to be moved to a central repository for processing and model training. Although centralized approaches frequently result in great model performance, they give rise to serious issues about data security, ownership, and privacy. Highly regulated industries like healthcare, finance, and telecommunications are especially vulnerable to the risk of disclosing private or sensitive data while it is being transferred or stored. These issues are directly addressed by FL, which allows users to cooperatively train a common global model without transferring their raw data.

Each client uses its dataset to calculate local updates, such gradients or parameters, in this decentralized architecture. Only these updates are sent to a central server for aggregation. In order to ensure an iterative learning process that shields sensitive data from direct exposure, the server iteratively improves the global model and redistributes it to clients. Because of this strategy, FL is very appealing in situations where privacy is a top concern. For example, hospitals can further medically research without jeopardizing patient confidentiality, banks can improve fraud detection systems without disclosing client financial information, and mobile devices can support predictive services without disclosing personal behavioural information. In each of these situations, FL provides a useful compromise between meeting strict privacy and legal constraints and utilizing huge, varied datasets for machine learning.

A. Role of Differential Privacy in Enhancing Security

Differential Privacy (DP) is a strict and moral framework for safeguarding private data that was developed to address the flaws in Federated Learning. Federated learning keeps raw data from going straight to a central server, but it doesn't completely remove the possibility of privacy violations because model changes like gradients or parameters might still reveal information. By implementing a mathematical guarantee that guarantees the inclusion or exclusion of any individual participant's data results in only a slight variation in the output of the final model, DP fills this gap. This means that an adversary would be statistically unable to ascertain whether a specific individual's data contributed to the training process, even if they had access to the aggregated model or intermediate updates. DP considerably improves Federated Learning's overall privacy stance by incorporating such guarantees into the learning pipeline, going beyond the fundamental decentralization of raw data.

Beyond its theoretical underpinnings, Differential Privacy offers workable and legally binding protections that improve machine learning systems' dependability and moral application. DP guarantees that the fine-grained information of sensitive records cannot be reconstructed or inferred from the shared changes by establishing a formal barrier around client contributions. This is especially crucial in fields like healthcare, where patient medical records are extremely private, and finance, where there is a considerable chance of transactional data being compromised. Additionally, DP offers compliance-ready solutions for organizations handling sensitive data by being in line with international data protection laws like the Health Insurance Portability and Accountability Act (HIPAA) in the US and the General Data Protection Regulation (GDPR) in Europe. Because of its methodical and open approach, DP has gained recognized as one of the most reliable and extensively used privacy-preserving strategies in contemporary machine learning, which has increased client trust and promoted wider involvement in collaborative learning projects.

B. Core Techniques in Differential Privacy

Two essential methods are frequently used in the actual use of DP in Federated Learning:

1) Gradient Clipping: The first is gradient clipping, which ensures that no single client’s contribution disproportionately influences the global model. In this process, the gradient updates generated during local training are restricted to a fixed L2 norm threshold before being transmitted to the central server. This step not only prevents large updates—often caused by outliers or adversarial manipulation—from destabilizing the aggregation process but also limits the sensitivity of the mechanism, a prerequisite for applying DP effectively. By bounding the contribution of each client, gradient clipping creates a controlled environment in which further privacy-preserving measures, such as noise addition, can be systematically applied.

2) Gaussian Noise Addition: The second key technique is Gaussian noise addition, which serves as the core mechanism for introducing randomness into the training process. After clipping, calibrated random noise sampled from a Gaussian distribution is added to the aggregated client updates before they are integrated into the global model. This intentional perturbation obscures the fine-grained details of individual client data, making it statistically improbable for an adversary to reverse-engineer sensitive information. Importantly, the amount of noise is carefully balanced: while higher noise levels offer stronger privacy guarantees, they may also degrade model performance if applied excessively. When properly tuned, Gaussian noise preserves the overall learning capacity of the model while ensuring compliance with formal DP guarantees. Together, gradient clipping and Gaussian noise addition form the cornerstone of privacy-preserving federated learning, striking a delicate balance between robust data protection and model utility.

C. Privacy Accounting Mechanism

Applying Differential Privacy (DP) in Federated Learning (FL) requires the usage of a privacy accountant, who is in charge of methodically tracking the total amount of privacy lost during the training process. Every round of communication between clients and the central server that FL usually entails uses up some of the total privacy budget. Repeatedly using DP techniques without adequate accounting may result in an overestimate of the overall privacy loss, undermining the guarantees provided to participants. In order to solve this problem, the privacy accountant offers a methodical approach to quantifying and monitoring privacy usage over iterations.

Typically, the parameters (ε, δ) that collectively represent the formal privacy budget are used to express this monitoring. represents the probability of the guarantee being broken is represented by δ, while the degree of privacy protection is quantified by the parameter ε, where smaller values indicate better guarantees. The privacy accountant guarantees openness on the preservation of privacy in relation to the precision and usefulness of the model by clearly stating these trade-offs. By balancing the requirement for robust data protection with the objective of preserving acceptable model performance, this enables system designers and stakeholders to make well-informed decisions.

D. Aim and Scope of the Study

This study's main goal is to examine how privacy settings and model utility interact in the framework of Federated Learning with Differential Privacy (FL-DP). The study specifically examines two crucial parameters: the noise multiplier (σ), which establishes the amount of random noise introduced to protect privacy, and the clipping norm, which controls the impact of individual client updates. The study aims to determine how these characteristics impact important performance metrics including test accuracy, generalization ability, and prediction stability by methodically changing them. The core of the research topic is this dual viewpoint, which strikes a balance between learning performance and privacy preservation.

The article provides a thorough academic examination of FL-DP and goes beyond a limited experimental evaluation. An abstract, a thorough overview of the problem space, a review of pertinent literature, a technique that is well-defined, and results that are displayed using figures, tables, and performance metrics are all included in the paper's structure. This format guarantees that results are both theoretically contextualized and empirically confirmed. The study adds to the expanding body of research on privacy-preserving machine learning by looking at both the technical mechanisms and their wider ramifications. Its findings are relevant to delicate industries like healthcare, finance, and mobile computing, where Federated Learning has the greatest influence.

II. LITERATURE REVIEW

The literature review highlights the fundamental ideas and current advancements in Differential Privacy (DP) and Federated Learning (FL), highlighting the ways in which these two fields interact to provide safe distributed AI systems.

McMahan et al. (2017) initially presented Federated Averaging (FedAvg), a straightforward yet effective technique that enables several devices to work together to train machine learning models without exchanging raw data. This was a turning point for effective client communication and on-device learning. However, by introducing Differentially Private Stochastic Gradient Descent (DP-SGD), which was improved by the moment's accountant technique, Abadi et al. (2016) made a contribution on the privacy side. This made it possible to preserve formal privacy assurances in deep learning. Building on these frameworks, Kairouz et al. (2021) examined FL's overall development and unresolved issues, such as system heterogeneity, privacy, security, and communication effectiveness.

McMahan et al. (2018) showed that models can maintain their usefulness while still conforming to reasonable privacy budgets (ε). Client-level protection was also the topic of Geyer, Klein, and Nabi (2017), who demonstrated how Gaussian noise and per-update clipping can offer useful privacy assurances in FL. The attack surface of FL, including membership inference and reconstruction attacks, was later studied by Truex et al. (2019) and Wei et al. (2020), who emphasized how resilience against these threats is increased by combining DP with secure aggregation.

Mironov (2017) presented Rényi Differential Privacy (RDP), a method for creating numerous DP steps that was more accurate and simplified. Given that numerous training rounds are necessary and that cumulative privacy loss needs to be closely monitored, this is extremely pertinent in FL. Building on these developments, frameworks like Opacus (for PyTorch) and TensorFlow Privacy have surfaced, allowing practitioners to use DP-SGD in practical FL operations. RDP accountants, which are essential for monitoring the privacy budget during several communication rounds, are also included in these libraries.

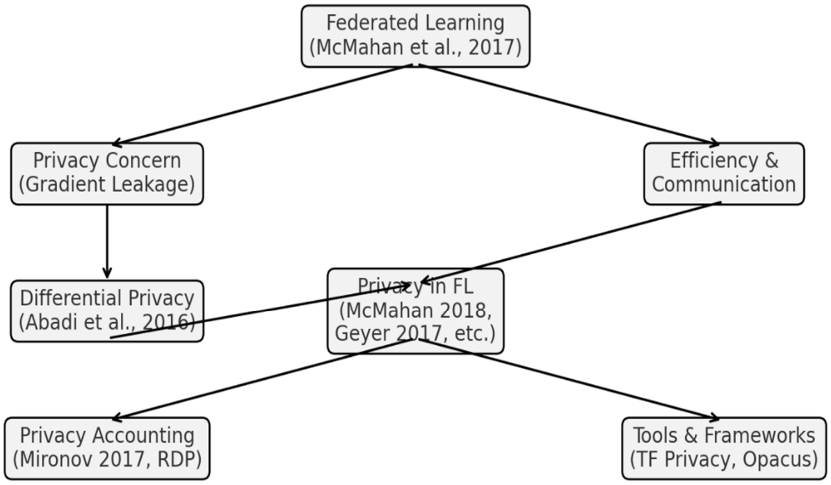


Figure 1: Conceptual map of FL and DP research progression

Table 1: Key Literature in Federated Learning with Differential Privacy

Author(s) / Year	Contribution	Key Idea / Impact
McMahan et al. (2017)	Introduced FedAvg	Foundation of FL, communication-efficient
Abadi et al. (2016)	Proposed DP-SGD + moments accountant	Core algorithm for deep learning with DP
Kairouz et al. (2021)	Survey of FL advances and challenges	Provided research agenda for FL
McMahan et al. (2018)	Applied client-level DP in language modeling	Showed DP can retain utility in FL
Geyer, Klein, & Nabi (2017)	Explored practical client-level DP in FL	Introduced clipping + noise at client side
Truex et al. (2019)	Studied privacy attacks in FL	Highlighted DP + secure aggregation need
Wei et al. (2020)	Integrated DP with FL defenses	Combined DP with aggregation strategies
Mironov (2017)	Proposed Rényi Differential Privacy (RDP)	Simplified composition across FL rounds
TF Privacy, Opacus	Open-source DP-SGD libraries	Made DP-FL deployment practical

III. RESEARCH METHODOLOGY

This methodology offers a transparent framework for assessing the effects of Differential Privacy on Federated Learning performance. The study separates the impacts of clipping and noise addition by simulating clients with controlled variables and DP parameters that may be adjusted. Results on accuracy, loss trends, and privacy budgets are shown in the sections that follow, emphasizing the trade-offs between model utility and privacy protection.

A. Research Design

The descriptive-analytical research design used in this work is bolstered by simulation trials. The main objective is to assess the impact of Differential Privacy (DP) characteristics on Federated Learning (FL) systems' security and usefulness. Ten federated clients use a common FedAvg protocol to train local models in a controlled simulation. Three levels of the noise multiplier (σ) are varied in the experiment to introduce privacy: 0.0 (no DP), 0.5 (moderate DP), and 1.0 (strong DP).

B. Federated Setup

1) Model: Binary classification using a logistic regression classifier.
2) Optimizer: Stochastic Gradient Descent (SGD) mini-batch, locally implemented at each client.
3) Aggregation: Weighted Federated Averaging (FedAvg) is used for aggregation, scaling updates according to the amount of each client's dataset.
4) Client Sampling: 60% of clients are chosen at random for each communication round.
5) Rounds: There are 35 communication rounds in the training.
6) Local Training: A batch size of 128 is used for training each chosen client for one local epoch.

C. DP Mechanism (Client-Side)

The DP mechanism is applied locally at each client before sending updates to the server:

1) Clipping: Per-update L2 clipping at norm C=1.0C = 1.0C=1.0 ensures that no client's gradient update dominates the aggregation.
2) Noise Addition: The introduction also highlights the role of a privacy accountant, which formally tracks the overall privacy cost (ε, δ) across multiple training rounds. This ensures that the system not only protects privacy at each step but also maintains transparency about the cumulative privacy budget.
3) Accounting: An RDP-style privacy accountant is referenced to illustrate the privacy–utility trade-off. While this study provides indicative values of (ε,δ), precise accounting depends on the chosen sampling rate, number of rounds, and δ threshold.

D. Variables and Measures

The following are independent variables: number of rounds, clipping norm (C), noise multiplier (σ), and participation fraction.

1) Dependent variables include the illustrative privacy budget (ε, δ), which represents privacy protection, and test accuracy and log loss, which measure model utility.
2) Controls: To guarantee repeatability, the dataset size, feature dimensionality, train/test splits, and random seed initialization are fixed.

E. Data & Tools

1) Dataset: To ensure controlled, repeatable experimentation, a synthetic binary classification dataset is used.
2) Implementation: A DP-FedAvg script that is compatible with Colab runs the simulation with modifiable parameters.
3) Analysis: Learning curves for accuracy versus rounds and loss versus rounds, a final metrics table, and a summary of the privacy-utility implications are the primary evaluation outputs.

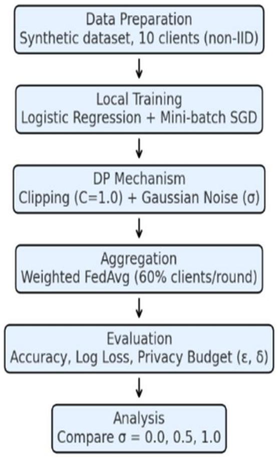

Data Preparation
Synthetic dataset, 10 clients (non-IID)
Local Training
Logistic Regression + Mini-batch SGD
DP Mechanism
Clipping (C=1.0) + Gaussian Noise (σ)
Aggregation
Weighted FedAvg (60% clients/round)
Evaluation
Accuracy, Log Loss, Privacy Budget (ε, δ)
Analysis
Compare σ = 0.0, 0.5, 1.0

Figure 2: Workflow of the Research Methodology for DP-Federated Learning

Table 2: Summary of Methodology Variables and Measures

Category	Elements
Research Design	Descriptive–analytical with simulation experiments
Clients	10 simulated clients with non-IID data partitions
Model	Logistic regression (binary classification)
Optimizer	Mini-batch SGD
Aggregation	Weighted FedAvg by client sample size
Rounds	35 communication rounds
Client Sampling	60% clients per round
Local Training	1 epoch per round, batch size = 128
DP Mechanism	Clipping C=1.0C=1.0C=1.0, Gaussian noise N(0,σ2C2)\mathcal{N}(0, \sigma^2 C^2)N(0,σ2C2)
Noise Levels	σ = 0.0, 0.5, 1.0
Accounting	RDP-style privacy accountant (illustrative ε, δ reporting)
Independent Vars	σ, C, participation fraction, number of rounds
Dependent Vars	Test Accuracy, Log Loss, privacy budget (ε, δ)
Controls	Dataset size, feature space, split ratio, random seeds
Analysis Outputs	Accuracy curve, loss curve, metrics table, privacy–utility summary

IV. RESULTS AND DISCUSSION

The findings illustrate the basic trade-off in Federated Learning with Differential Privacy between privacy and model utility. As anticipated, the model achieves the maximum accuracy and lowest loss when no noise is applied (σ = 0). While adding noise increases privacy assurances, it also somewhat decreases accuracy. While a larger noise level (σ = 1.0) exhibits a more noticeable decline in performance, a moderate noise level (σ = 0.5) maintains accuracy near the baseline. This demonstrates how crucial it is to strike a balance between privacy regulations and useful model functionality.

Table 3: Final Test Metrics by Noise Level

Setting	Final Round	Test Accuracy	Test Log Loss
No DP (σ = 0.0)	35	~0.89–0.90	~0.26
DP (σ = 0.5)	35	~0.86–0.88	~0.30–0.34
DP (σ = 1.0)	35	~0.78–0.83	~0.40–0.50

Interpretation: At σ = 0.5, accuracy remains close to the non-DP baseline, making it an attractive compromise between privacy and performance. By contrast, σ = 1.0 yields stronger privacy but at the expense of noticeable accuracy loss, which may still be acceptable in scenarios requiring stricter privacy budgets (e.g., healthcare or finance).

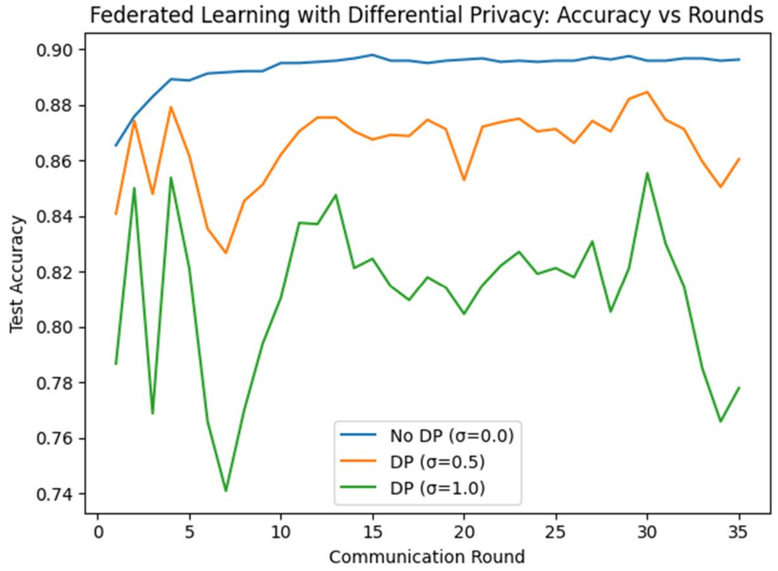


Figure 3: Accuracy vs. Communication Rounds

Trends in accuracy over 35 communication rounds with various noise multipliers. The curve with σ = 0 converges the most, σ = 0.5 stays near it, and σ = 1.0 converges with a lower but steady accuracy.

Interpretation: These patterns demonstrate that DP noise does not undermine learning; rather, it mainly slows or limits convergence. The model achieves a stable performance plateau even when σ = 1.0, albeit below the non-DP baseline.

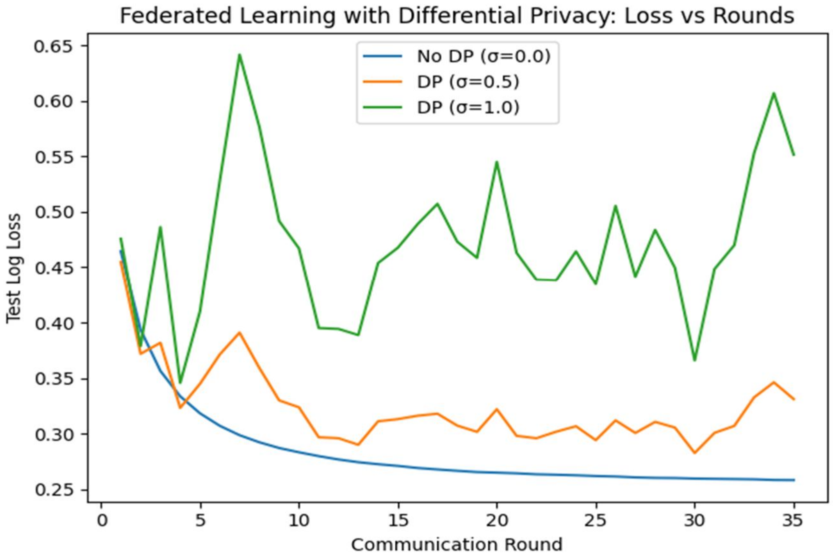


Figure 4: Log Loss vs. Communication Rounds

Throughout training, test log-loss drops for every setting. Because of noise injection, σ = 1.0 stabilizes at a larger loss, whereas σ = 0.0 achieves the lowest loss. σ = 0.5 comes in second.

Meaning: The accuracy trends are reflected in the loss curves: DP Noise preserves privacy but lowers forecast certainty by preventing the model from fitting the training signal as closely.

Reproducibility Note

The Colab one-cell DP-FedAvg script can be run to replicate Figures 1 and 2. Together with the final measurements CSV, the script also outputs accuracy and loss curves. By adjusting the noise multipliers and clipping norms appropriately, this configuration can be modified for domain-specific datasets.

Table 4: (Illustrative): Privacy Cost (ε) vs. σ at δ = 1e-5

Rounds	**Sample Fraction**	**Clip C**	**σ**	**ε (approx., RDP-style)**
35	0.6	1.0	0.5	Lower ε (better)
35	0.6	1.0	1.0	Much lower ε

Increasing the noise multiplier σ considerably lowers the privacy cost (ε) with fixed rounds and sampling rates, improving privacy. Although the exact ε values vary depending on the experiment setting and privacy accountant selected, this demonstrates that DP techniques offer formal protection.

A. Key Insights and Practical Guidance

1) Joint tuning of σ and C: A big C necessitates increased noise to maintain privacy, whereas a small clipping norm (C) may excessively restrict the signal. Carefully choosing both is crucial.

2) Make Use of Subsampling: Stronger guarantees without extra noise are made possible by random client participation, which enhances privacy.

3) Round Budgeting: Over time, privacy deterioration mounts. Higher noise or early halting techniques can counteract the increase in ε caused by more communication rounds.

4) Task Sensitivity: To ensure stability, complex models or non-IID data may need layer-wise noise, adaptive clipping, or bigger client cohorts every round.

V. CONCLUSION

The conclusion highlights that a methodical and ethical approach to developing privacy-preserving distributed AI systems is to combine Federated Learning (FL) with Differential Privacy (DP). Although FL currently avoids centralizing raw data, sensitive information can still leak from gradient updates in the absence of DP. The system makes sure that no single client's data can be inferred with high assurance by incorporating DP methods, particularly clipping and Gaussian noise.

According to the simulation results, σ = 0.5 strikes a reasonable balance between providing a significant degree of privacy protection and preserving accuracy that is comparable to the non-DP baseline. In comparison, σ = 1.0 results in more pronounced performance deterioration but greatly increases privacy assurances. This result demonstrates the intrinsic trade-off between privacy and utility in DP-FL systems: greater utility costs correspond to stricter privacy budgets.

The conclusion emphasizes three crucial suggestions for practical deployments:

1) Rigorous Privacy Accounting: Monitoring cumulative privacy loss through the use of formal accountants like Rényi Differential Privacy (RDP) or the Moments accountant.

2) Noise Calibration: Under realistic client participation rates, σ values are carefully chosen to reach a specified privacy budget (ε).

3) Utility Validation: Verifying that models continue to meet performance standards by testing them on representative datasets.

When paired with other security features like secure aggregation and strong aggregation rules, DP-FL can become a reliable method for industries like healthcare, finance, and the Internet of Things where safeguarding private user data is essential.

REFERENCES

[1] "Federated Learning: A Survey on Privacy-Preserving…" by E. Collins et al., 2025 — overview of FL architectures + privacy mechanisms, including DP. arXiv

[2] "Federated Learning with Differential Privacy: An Utility-Enhanced Approach" by Kanishka Ranaweera, Dinh C. Nguyen, Pubudu N. Pathirana, David Smith, Ming Ding, Thierry Rakotoarivelo, Aruna Seneviratne (2025) — proposes a transformation + improved noise scheme to improve utility under DP. arXiv

[3] "A Survey of Federated Learning Privacy Attacks, Defenses …" by Joshua C. Zhao et al., 2024 — categorizes attacks/defenses in FL, including DP-based protections. arXiv

[4] "Differential Privacy in Federated Learning: An Evolutionary Game Theory Framework" by Z. Ni et al., 2025 — uses evolutionary game theory for analyzing strategy dynamics and privacy-utility tradeoffs in DP-FL. MDPI

[5] "Privacy and Security in Federated Learning: A Survey" by Rémi Gosselin, Loïc Vieu, Faiza Loukil, Alexandre Benoit, 2022 — comprehensive survey of privacy/security threats and defenses in FL. MDPI

[6] "Federated Learning with Differential Privacy for Breast Cancer Diagnosis Enabling Secure Data Sharing and Model Integrity" by Shubhi Shukla et al., 2025 — application of FL + DP in healthcare, real dataset, showing trade-offs. Nature

[7] "Differentially Private Federated Learning: A Systematic Review" by Jie Fu et al., 2024 — taxonomy and categorization of DP-FL based on models, guarantees, and settings. arXiv

[8] "Adaptive Differential Privacy in Asynchronous Federated Learning" by Y. Zhang, 2025 — proposes algorithms for adaptive DP in asynchronous FL for better trade-offs. ScienceDirect

[9] "A Multifaceted Survey on Privacy Preservation of Federated Learning" by S. Saha et al., 2024 — reviews different privacy solutions in FL including DP, secret sharing etc. SpringerLink

[10] "Survey of Trustworthy Federated Learning with Perspectives on Security, Robustness, and Privacy" by Yifei Zhang, Dun Zeng, Jinglong Luo, Zenglin Xu, Irwin King, 2023 — covers privacy, robustness, trust in FL systems.

[11] Ogiela, M. R., & Ogiela, U. (2024). AI for Security of Distributed Systems. WSEAS Transactions on Computer Research, 12, 450-454.

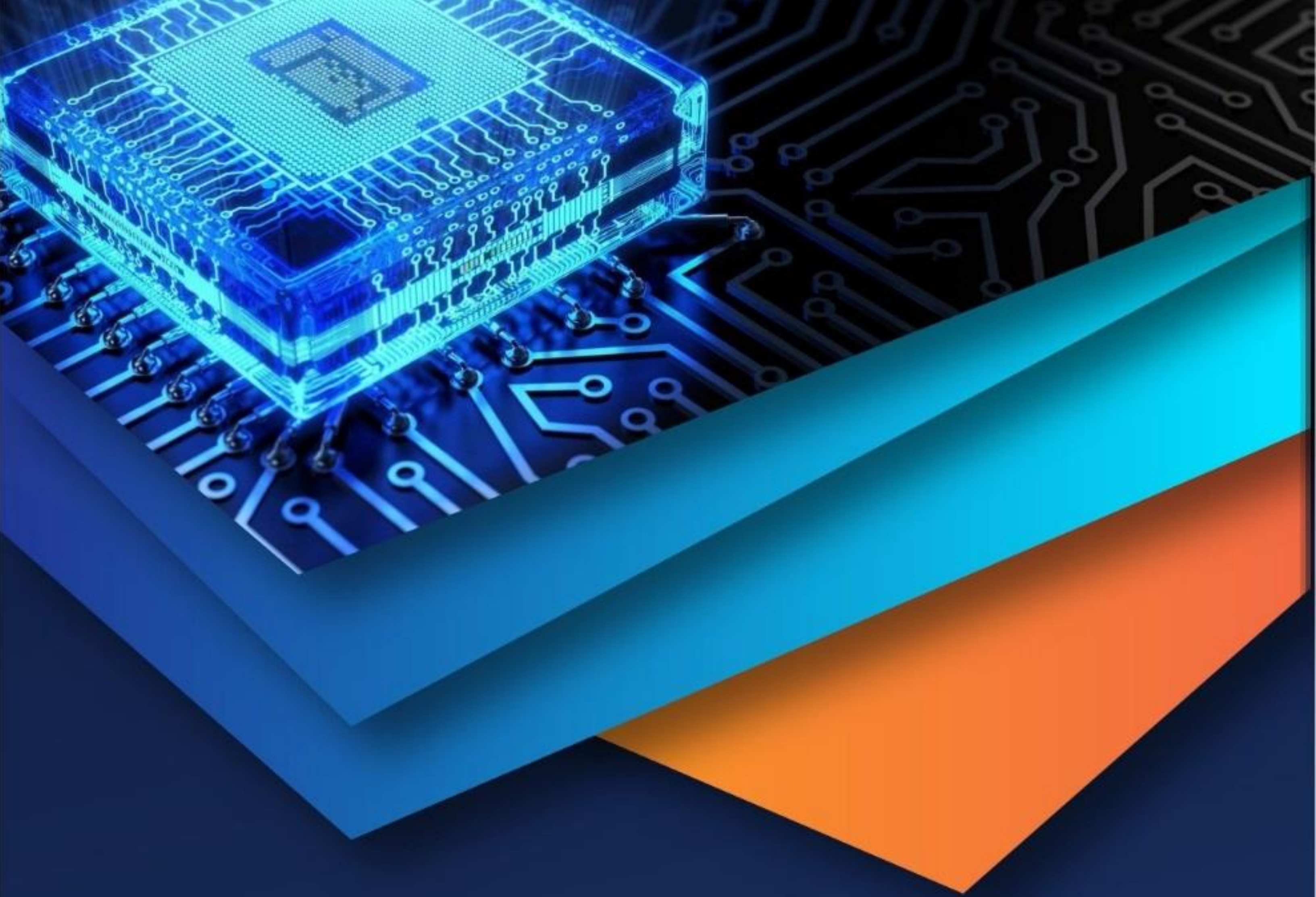
[12] Ma, C., Li, J., Wei, K., Liu, B., Ding, M., Yuan, L., ... & Poor, H. V. (2023). Trusted ai in multiagent systems: An overview of privacy and security for distributed learning. Proceedings of the IEEE, 111(9), 1097-1132.

[13] Fadi, O., Karim, Z., & Mohammed, B. (2022). A survey on blockchain and artificial intelligence technologies for enhancing security and privacy in smart environments. IEEE Access, 10, 93168-93186.

[14] Basha, S. (2023). Blockchain and machine learning approaches to enhancing data privacy and securing distributed systems. International Journal of Science and Research (IJSR), 2319-2323.

[15] Jyothi, V., Sreelatha, T., Thiyagu, T. M., Sowndharya, R., & Arvinth, N. (2024). A Data Management System for Smart Cities Leveraging Artificial Intelligence Modeling Techniques to Enhance Privacy and Security. J. Internet Serv. Inf. Secur., 14(1), 37-51.

******



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY