



SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Ensemblefedshield-Pruda: A Privacy-Preserving Adaptive Federated Learning Framework with Relevance- and Uncertainty-Guided Dynamic Update Attenuation for Mental Health Risk Prediction

M. Malathi¹, Dr. M. RameshKumar²

¹Research Scholar, PG & Research Department of Computer Science, Government Arts College, Nandanam, Chennai 600035, Tamil Nadu, India. E-mail: malathimano2021@gmail.com

²Associate Professor and Head, PG & Research Department of Computer Science, Government Arts College, Nandanam, Chennai 600035, Tamil Nadu, India. E-mail: proframeshkumar@gmail.com

ABSTRACT

Federated Learning (FL) is becoming more popular for mental health risk prediction because it enables collaborative model training while keeping sensitive patient data localised and avoiding direct data exchange. Yet the distribution of model updates can still lead to the leakage of private information via gradient leakage, membership inference, and model inversion attacks. On the other hand, traditional Differential Privacy (DP) approaches generally result in a considerable loss of utility because of the noise they add and the use of fixed clipping methods. In order to overcome these disadvantages, A Privacy-Preserving Adaptive Federated Learning Framework with Relevance and Uncertainty-Guided Dynamic Update Attenuation (ENSEMBLEFEDSHIELD-PRUDA) is a framework that has been developed for safe and secure mental health risk prediction and offers protection against various threats. This framework manages the training of a Multi-Layer Perceptron (MLP) by means of moving-average adaptive clipping, gradient-based feature-relevance scoring, Monte Carlo dropout for uncertainty estimation, uncertainty-guided dynamic attenuation, and round-based adaptive differential privacy. These mechanisms ensure that the updates are protected before the secure global aggregation takes place, preserve useful model parameters, reduce unreliable client updates, and stabilise local optimisation. The effective performance of the proposed framework is further verified by extensive comparative experiments with five baseline models, namely Centralised MLP, Standard Federated Learning (FedAvg), and Federated Learning with Fixed Differential Privacy, Federated Learning with Adaptive Clipping and Fixed DP, FedAvg with Uncertainty Guided DP, all of which show consistent improvements in predictive performance while still lacking strong privacy guarantees. As a result, the proposed framework attains a better balance between privacy, utility and enhances its resistance to information leakage. Finally, the experimental results indicate that the proposed framework (ENSEMBLEFEDSHIELD-PRUDA) offers robust privacy preservation and reliable mental health risk prediction, which means it is suitable for secure collaborative healthcare analytics.

Keywords: Federated Learning, Differential Privacy, Mental Health Risk Prediction, Privacy-Aware Learning, Adaptive Clipping, Monte Carlo Dropout, Uncertainty Estimation, Gradient Attenuation, Secure Healthcare Analytics.

I. Introduction

The rapid adoption of intelligent mental health care systems has made it possible to develop data-driven decision support models that are able to detect depression, anxiety, stress, and other psychological disorders from patient data. Data privacy is one of the most significant issues in healthcare analytics given that these applications involve the use of very sensitive personal information. Traditional centralised machine learning involves collecting patient data on a single server, which increases the risk of unauthorised access, breaches of privacy, and failure to comply with regulations. Federated Learning (FL) is a decentralised machine learning approach that allows a global model to be trained collaboratively while keeping the patient-sensitive data on the client devices themselves. Even though the raw data never leave the client device, the model updates sent over the network can still be exposed through attacks such as model reconstruction, membership inference, or gradient inversion. Differential Privacy (DP) is generally employed to add calibrated noise to the client updates before they are aggregated. However, conventional DP methods usually apply a fixed amount of noise and use constant clipping thresholds, and therefore are not able to adapt to the changing features of the client updates during training. These limitations often lead to a decrease in model performance while unnecessarily disrupting informative updates. In order to address these limitations, this research paper presents a (ENSEMBLEFEDSHIELD-PRUDA) framework for secure and protected federated mental healthcare analytics.

The framework presented includes a number of complementary adaptive techniques that together enhance both privacy preservation and the reliability of learning. At the beginning, it uses moving-average adaptive clipping to dynamically determine clipping thresholds by referring to the historical norms of the updates, which allows for stable optimisation while ensuring that large client updates do not take control of the global learning process. Moreover,

uncertainty estimation is obtained using Monte Carlo Dropout to measure the confidence level of each local client model's predictions. The suggested Update Relevance Scoring mechanism evaluates the relative importance of the client updates by means of a gradient-based analysis, instead of giving the same degree of protection to every parameter. By doing this, it is possible to identify those updates that are highly informative before carrying out privacy preservation. These uncertainty estimates are then incorporated into an Uncertainty-Guided Dynamic Update Attenuation method, which reduces faulty client updates while at the same time keeping trustworthy information during the process of collaborative optimisation.

Finally, Round-Based Adaptive Differential Privacy introduces communication-aware Gaussian noise whose magnitude is varied according to the progress of training, thus offering strong privacy protection without unreasonably lowering model accuracy. The framework is tested on a mental health risk prediction dataset within a federated learning environment using a Privacy-Aware Multilayer Perceptron (MLP) classifier. An experimental evaluation demonstrates that the proposed adaptive optimisation strategy improves the balance between predictive performance and privacy preservation and also increases robustness against a range of client-side and server-side security threats. The combination of update relevance scoring, uncertainty-guided attenuation, adaptive differential privacy, and secure aggregation provides a thorough solution for reliable and privacy-preserving collaborative mental healthcare analytics. Unlike conventional federated learning methods that treat privacy and security separately, the proposed framework jointly addresses privacy preservation, adaptive optimisation, and the privacy-utility trade-off. The following five baseline models are used for comparison in order to evaluate the performance of the proposed framework:

(i) a Centralized MLP model that does not use Federated Learning, (ii) a Standard Federated Learning model based on FedAvg, (iii) Federated Learning model that incorporates Fixed Differential Privacy, (iv) Federated Learning with Adaptive clipping with Fixed DP, and (v) FedAvg with Uncertainty Guided DP. These baseline models serve as a systematic benchmark against which the contributions of the proposed adaptive framework can be assessed by progressively enhancing both privacy preservation and distributed learning capabilities. The comparative analysis shows how well adaptive clipping, round-based adaptive differential privacy, uncertainty-guided update attenuation, and update relevance scoring can be incorporated into a single Federated learning architecture.

The major objectives of this work are mentioned as follows:

- ❖ To develop the ENSEMBLEFEDSHIELD-PRUDA framework for privacy-preserving adaptive federated learning in mental-health risk prediction.
- ❖ To develop a Moving Average Adaptive Clipping strategy for dynamic regulation of client update magnitudes during federated optimization.
- ❖ To design a gradient-based Update Relevance Scoring mechanism for identifying informative client updates.
- ❖ To regulate unreliable client updates by combining Monte Carlo Dropout-based uncertainty estimation with Uncertainty-Guided Dynamic Update Attenuation.
- ❖ To develop a Round-Based Adaptive Differential Privacy mechanism for dynamically adjusting Gaussian noise across communication rounds while preserving model utility.

II. Related Literature

Privacy-preserving Federated Learning (FL) has emerged as a successful paradigm for collaborative machine learning because it allows for distributed model training without sharing sensitive raw data. Recent surveys have comprehensively examined major privacy threats in FL, such as gradient leakage, inference attacks, model inversion, poisoning attacks, and communication vulnerabilities, while examining countermeasures such as Differential Privacy (DP), Secure Multi-Party Computation (SMC), homomorphic encryption, and secure aggregation [1], [2], [6]. By combining decentralized learning with differential privacy and transfer learning, these frameworks significantly minimize the exposure of sensitive patient data while facilitating collaborative model development across distributed healthcare institutions. Despite these developments, major existing approaches primarily concentrate on privacy preservation and predictive performance while providing limited consideration for the adaptive reliability of client updates during federated optimization [11], [8].

Differential privacy has become one of the most popular privacy-preserving techniques in federated learning, and it generally injects adaptive noise and adaptive clipping at different clients and rounds to satisfy formal (ϵ, δ) -differential privacy guarantees. More recently, adaptive privacy mechanisms have dynamically adjusted clipping thresholds and noise levels to improve the privacy-utility trade-off. Communication-efficient frameworks such as CLFLDP and adaptive personalized federated learning further improve optimization through layer-wise clipping, adaptive privacy budget allocation, and dynamically sparsified client updates [12].

Another important research direction focuses on preventing information leakage and handling threats via gradient leakage. Gradient leakage studies demonstrate that exchanged gradients may unintentionally reveal sensitive training samples even when raw client data remain locally stored [3]. Similarly, privacy-preserving federated learning frameworks developed for the Industrial Internet of Things (IIoT), natural language processing, and smart healthcare integrate secure aggregation, local differential privacy, and communication-efficient optimization techniques to strengthen communication security [4], [5], [6]. Although these approaches significantly enhance privacy protection,

they generally treat all client updates equally during optimization, without differentiating highly informative updates from less informative ones. Prediction uncertainty has recently become an important research topic in trustworthy federated learning. Bayesian uncertainty estimation methods, particularly entropy or Monte Carlo dropout, have been successfully adopted for clinical decision support and the analysis of the performance with and without differential privacy with non-IID data and the integration of differential privacy with personalized federated learning using non-IID data for client updates. [12][13][14]. However, uncertainty information is rarely incorporated into adaptive privacy-preserving optimization. Consequently, uncertain client updates continue to participate in collaborative optimization with equal importance, potentially reducing global model robustness and convergence performance [2], [4].

Overall, existing studies have independently investigated adaptive clipping, differential privacy, uncertainty estimation, communication-efficient federated learning, and privacy-preserving healthcare applications [1]–[15]. However, three important research gaps remain. First, existing federated learning frameworks rarely evaluate the relevance of client updates before applying differential privacy, resulting in uniform perturbation of both informative and less informative updates. Second, predictive uncertainty is rarely incorporated into adaptive client update regulation, allowing unreliable local models to influence global optimization equally. Third, adaptive clipping, update relevance evaluation, uncertainty estimation, dynamic update attenuation, and adaptive differential privacy are generally investigated as independent techniques rather than being integrated into a unified optimization framework. Motivated by these limitations, the proposed Privacy-Aware Update Relevance Scoring with Uncertainty-Guided Dynamic Update Attenuation Adaptive Differential Privacy (PAURS-UGDUA-ADP) framework integrates Moving Average Adaptive Clipping, Update Relevance Scoring, Monte Carlo Dropout-based uncertainty estimation, Dynamic Update Attenuation, and Round-Based Adaptive Differential Privacy into a unified federated learning architecture. By jointly exploiting update relevance and predictive confidence before adaptive privacy preservation, the proposed framework improves optimization stability, preserves informative client updates, achieves a better privacy-utility trade-off, and enhances privacy-preserving mental health prediction.

III. Methodology

This ENSEMBLEFEDSHIELD-PRUDA framework works on a mental health risk dataset, containing 25,000 patient records in a federated learning setting. Initially, a Multi-Layer Perceptron (MLP) trains the model, and then moving-average adaptive clipping regulates gradient magnitudes; a gradient-based update-relevance scoring mechanism is used, followed by Monte Carlo (MC) estimation to predict uncertainty. These two mechanisms jointly guide dynamic update attenuation, which reduces unreliable updates before privacy preservation. Finally, round-based adaptive noise differential privacy adds Gaussian noise to the communication stage, and the privacy-preserved updates are aggregated using Federated Averaging (FedAvg) to develop a global model.

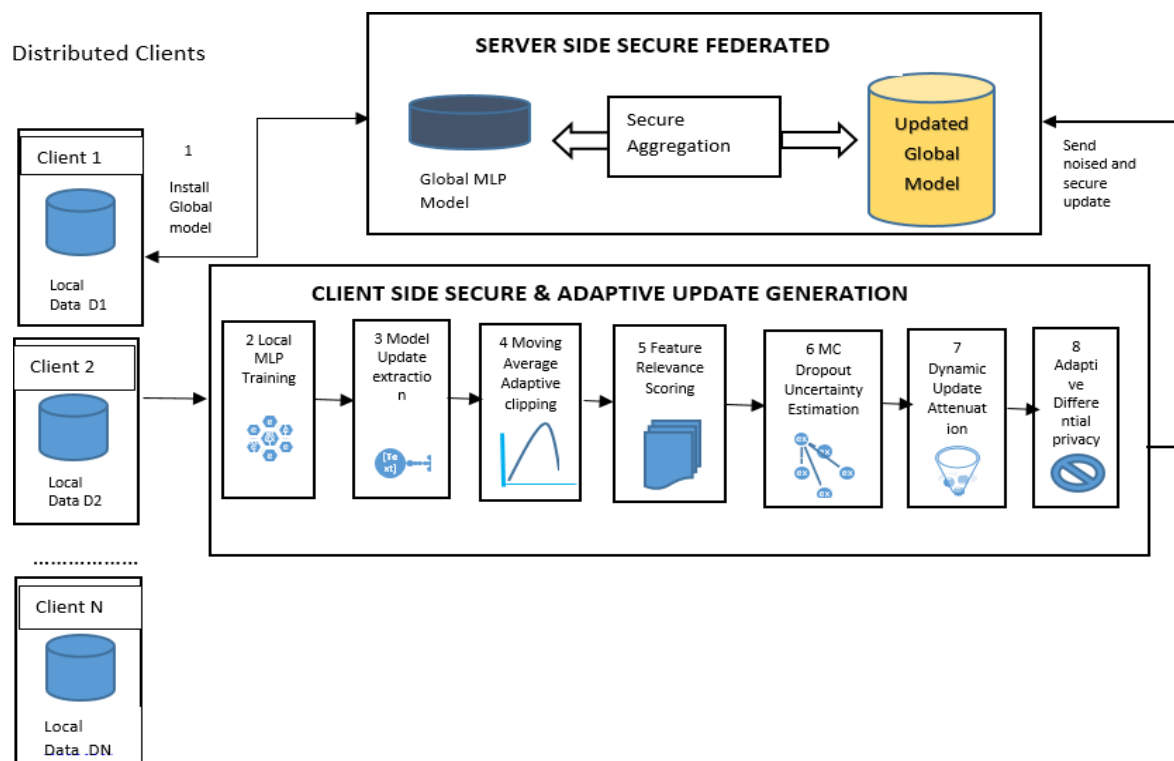


Figure 1: Architecture flow diagram of the proposed system.

Figure 1 demonstrates the complete architectural workflow of the proposed ENSEM-BLEFEDSHIELD-PRUDA framework. This framework enables local training on each client using a Privacy-Aware MLP, applying adaptive moving-average clipping, feature update-relevance analysis, and MC dropout uncertainty estimation, followed by Dynamic update attenuation and adaptive noise. Finally, the threat prevented secure client updates from being sent to the updated global model. This framework prevents client-side and server-side threats such as Gradient leak-age, model inversion, membership inference, Reconstruction of Private Information, Privacy Leakage during Aggregation, etc. The Federated Learning workflow below demonstrates how local model training; moving-average adaptive clipping; update-relevance scoring; MC dropout-based uncertainty estimation; uncertainty-guided dynamic update attenuation; round-based adaptive differential privacy; and federated model aggregation are performed in this proposed framework.

3.1. Local Privacy Preserving Model Training

The training dataset is distributed among several federated clients, so that sensitive patient records are never stored anywhere other than locally during the learning process. Each client pre-processes its local data by standardising features and trains a Privacy-Aware Multi-Layer Perceptron with three hidden layers, Batch Normalisation, ReLU activation, and Dropout regularisation. The model parameters are optimised through the Adam optimiser with cross-entropy loss. The

objective function for local optimisation is: $L = \frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i)$, where y_i represents the true

class label and $\hat{y}_i$ denotes the predicted probability.

3.2. Moving Average Adaptive Clipping

Instead of using a fixed clipping threshold, the proposed work dynamically adjusts the clipping threshold to the moving average of previous client update norms. This adaptive strategy stabilises optimisation by securing excessively large model updates while avoiding unnecessary clipping of informative gradients. Consequently, model sensitivity remains bounded before applying differential privacy. The adaptive clipping threshold is computed as: $C_t = \beta C_{t-1} + (1 - \beta) \|g_t\|$, where C_t is the clipping threshold, β denotes the momentum coefficient, and $\|g_t\|$ represents the current update norm.

3.3. Update Relevance Scoring

After moving adaptive clipping, the importance of each model update is estimated using gradient-based relevance analysis. The average absolute gradient magnitude is calculated to identify highly informative updates that contribute to the learning process. Unlike traditional methods that perturb all updates equally, the proposed method preserves important updates while attenuating less informative updates more strongly before privacy preservation. The relevance score is

calculated as $R = \frac{1}{N} \sum_{i=1}^N \|\nabla x_i\|$, where ∇x_i denotes the input gradient corresponding to feature x_i .

3.4. MC Dropout-Based Uncertainty Estimation

Estimate prediction reliability, the proposed method governs Monte Carlo Dropout during inference. Multiple stochastic forward passes are performed while keeping dropout layers active, producing an approximate Bayesian uncertainty estimate for each client model. The resulting uncertainty values are subsequently used to regulate the contribution of client updates during optimisation. Predictive uncertainty is estimated as: $U = \text{Var}(P_1, P_2, \dots, P_T)$, where P_t denotes the predicted probability obtained from the t^{th} stochastic forward pass.

3.5. Uncertainty-Guided Dynamic Update Attenuation:

The proposed framework integrates Update Relevance Scores and predictive uncertainty to calculate an adaptive attenuation factor for each client update. Updates with higher uncertainty receive stronger attenuation, whereas highly relevant updates retain greater influence during optimisation. This adaptive mechanism reduces the impact of unreliable updates while preserving informative model parameters. The attenuation factor is computed as: $A = \lambda(1 - U) + (1 - \lambda)R$, where U represents predictive uncertainty, R denotes the relevance score, and λ controls the contribution of both components.

3.6. Round-Based Adaptive Differential Privacy:

After applying update attenuation, adaptive Gaussian noise is inserted into every client update to satisfy differential privacy. Unlike conventional methods that use fixed noise throughout training, the proposed framework adjusts the privacy noise by communication round, improving the privacy-utility trade-off while maintaining formal privacy guarantees. The privacy-preserving update is generated as: $W = W + N(0, \sigma^2)$, where W is the attenuated update and

σ represents the adaptive Gaussian noise scale.

3.7. Federated Model Aggregation:

The privacy-preserved client updates are transmitted to the central server, where Federated Averaging aggregates the received model parameters to develop the global model. Only model parameters are exchanged during communication, while all patient data remain locally saved on individual clients, thereby securing data confidentiality throughout collaborative learning. The

global model is updated using $W^{t+1} = \sum_{k=1}^K \frac{n_k}{N} W^t$, where W^t denotes the local model of client k ,

n_k represents the number of local samples, and N is the total number of training samples across all participating clients. The following Algorithm represents how this proposed framework of EN-SEMBLEFEDSHIELD-PRUDA works in the mental health risk dataset.

Algorithm 1: Proposed ENSEMBLEFEDSHIELD-PRUDA Framework

Algorithm 1:ENSEBLEFEDSHIELDPRUDA

Input:

Federated client datasets $D_1, D_2, \dots, D_k$

Output:

Trained global model W

- 1: *Partition the training data among K clients*
- 2: *Initialize the global Privacy-Aware MLP model W*
- 3: *for each communication round t do*
- 4: *for each client $k = 1, 2, \dots, K$ do*
- 5: *Train the local MLP using the local dataset D_k*
- 6: *Compute the local model update ΔW_k*
- 7: *Apply Moving Average Adaptive Clipping to ΔW_k*
- 8: *Compute the Update Relevance Score*
- 9: *Estimate uncertainty using MC Dropout*
- 10: *Perform Dynamic Update Attenuation*
- 11: *Inject Round-Based Adaptive DP Gaussian Noise*
- 12: *end for*
- 13: *Aggregate the client updates using FedAvg*
- 14: *Update the global model W*
- 15: *Check the convergence condition*
- 16: *end for*
- 17: *Return the trained global model W*

IV. Experimental Setup

The ENSEMBLEFEDSHIELD-PRUDA framework looked at a mental health risk dataset which included 25000 samples and 24 predictive features relating to demographic, behavioural, lifestyle, and psychological factors. Prior to beginning the model training, data pre-processing—designed to clean the data—was carried out on the dataset to deal with duplicate records and miss-ing values; the categorical attributes were label encoded and the numerical features were stand-ardised with the use of StandardScaler. The dataset was then randomly divided into 80% for train-ing and 20% for testing, with the class distribution being kept unchanged by means of stratified sampling.

The training data was divided into three non-overlapping subsets in order to mimic a federated learning situation involving three clients. Each client trained a Privacy-Aware Multi-Layer Per-cep-tron (MLP) which had three fully connected hidden layers, used ReLU activation, included batch normalisation and employed dropout as a regularisation technique. The Adam optimiser was employed to train the model with a batch size of 64, five local epochs per communication round, a learning rate of 0.001, and a weight decay of 1×10^{-4} . The federated model was trained for ten communication rounds and the global model parameters were updated by means of the Federated Averaging (FedAvg) algorithm. The adaptive privacy-preserving architecture that is proposed consists of Moving Average Adaptive Clipping, Update Relevance Scoring, MC Dropout-based Uncertainty Estimation, Dynamic Update

Attenuation, and Round-Based Adaptive Differential Privacy prior to model aggregation. Differential privacy was implemented using the Gaussian method with a privacy budget of $\epsilon = 1.0$ and $\delta = 1e-5$. The injected noise was adaptively tuned according to the communication round. The table below shows the parameters and their values used in this proposed framework.

Table 1: Experimental Parameters with Values of ENSEMBLEFEDSHIELD-PRUDA framework

Parameter	Value
Dataset	Mental Health Risk Dataset
Total Samples	25,000
Features	24
Target Classes	3
Train/Test Split	80% / 20%
Number of Clients	3
Global Model	Privacy-Aware MLP
Hidden Layers	256-128-64
Activation	ReLU
Batch Normalization	Yes
Dropout	0.20
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Local Epochs	5
Communication Rounds	10
Aggregation	FedAvg
Differential Privacy	Gaussian Mechanism
Privacy Budget (ϵ)	1.0
Delta (δ)	1e-5
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, ROC-AUC

V. Result And Discussion

The proposed A Privacy-Preserving Adaptive Federated Learning Framework With Relevance- And Uncertainty-Guided Dynamic Update Attenuation (ENSEMBLEFEDSHIELD-PRUDA) framework was assessed using the mental health dataset in a federated learning environment with three distributed clients. The proposed architecture dynamically adjusts model updates by integrating Moving Average Adaptive Clipping, Feature Relevance Scoring, MC Dropout-based Uncertainty Estimation, Dynamic Update Attenuation, and Round-Based Adaptive Gaussian Differential Privacy; it overcomes the disadvantages of fixed clipping and constant differential privacy noise in conventional federated learning approaches. This proposed framework is compared with baseline models such as (i) Centralised MLP, (ii) Standard federated learning (FedAvg), and (iii) Federated Learning with fixed differential privacy, (iv) Federated Learning with Adaptive clipping and Fixed DP, and (v) FedAvg with Uncertainty Guided DP, the results of the proposed framework preserves informative updates: these adaptive components stabilise local optimisation, suppress uncertain model parameters, and regulate privacy noise with respect to the training progress. In this approach, adaptive clipping minimises excessive gradient distortion, feature-aware update modulation secures highly informative model parameters, and uncertainty-guided attenuation prevents unreliable updates from dominating global optimisation. Experimental results show that combining adaptive optimisation with privacy preservation consistently improves predictive performance while maintaining rigorous privacy guarantees. Overall, these findings show that integrating adaptive optimisation strategies with differential privacy significantly improves federated learning performance for mental healthcare predictions. It also achieves an effective privacy-utility trade-off while maintaining stable convergence, making it suitable for privacy-sensitive distributed healthcare applications. The tables below demonstrate the overall baseline performance comparison for the proposed ENSEMBLEFEDSHIELD-PRUDA framework.

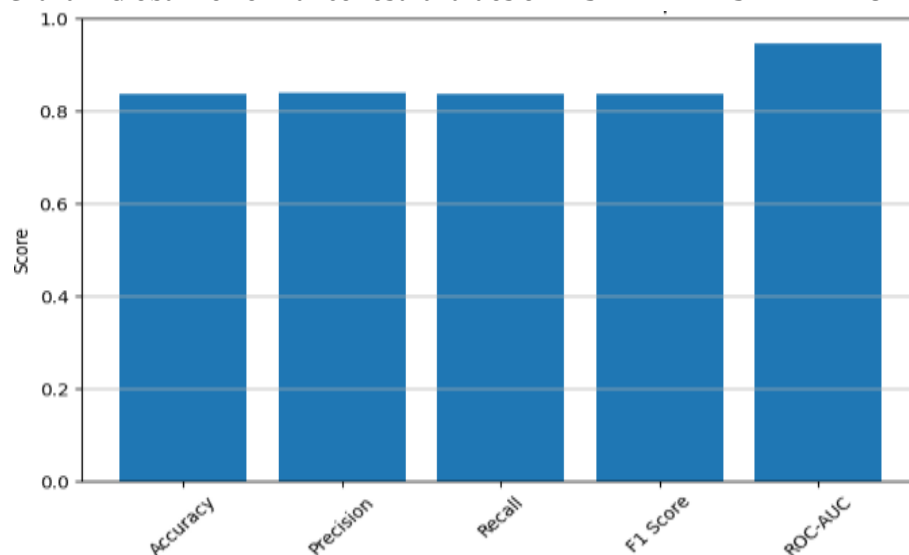
Table 2: Overall Performance Comparison

Method	Accuracy	Precision	Recall	F1-score	ROC-AUC
Centralized MLP	0.7942	0.7960	0.794	0.7939	0.9174

Standard FL (FedAvg)	0.7472	0.7478	0.7472	0.7464	0.8760
FL + Fixed DP	0.7422	0.7508	0.742	0.7310	0.8797
Federated Learning + Adaptive Clipping + Fixed DP	0.7638	0.7835	0.763	0.7544	0.9002
Fedavg + Uncertainty-Guided DP	0.6178	0.6718	0.6178	0.6190	0.7859
Proposed ENSEMBLEFEDSHIELD-PRUDA	0.8364	0.8389	0.8364	0.8367	0.9466

Table 2 show that various baseline models, such as Centralised MLP, Standard FL, Federated learning with fixed differential privacy, Federated learning with adaptive clipping and fixed differential privacy, and FedAvg with uncertainty-guided DP, are analysed and evaluated with the same mental health risk dataset using metrics like accuracy, precision, recall, F1 score, and ROC AUC. These findings show that the proposed ENSEMBLEFEDSHIELD-PRUDA framework achieves higher accuracy, precision, recall, F1 score, and AUC. It clearly shows that the proposed framework helps to protect patients' records in the dataset, prevent threats, secure model updates, and improve the privacy-utility trade-off.

Chart 1: Global Performance result values of ENSEMBLEFEDSHIELD-PRUDA



The Chart 1 represents the overall results obtained from the proposed framework.

CONFUSION MATRIX				
[[1533 337 1]				
[188 2032 145]				
[0 147 617]]				
CLASSIFICATION REPORT				
	precision	recall	f1-score	support
0	0.89	0.82	0.85	1871
1	0.81	0.86	0.83	2365
2	0.81	0.81	0.81	764
accuracy			0.84	5000
macro avg	0.84	0.83	0.83	5000
weighted avg	0.84	0.84	0.84	5000

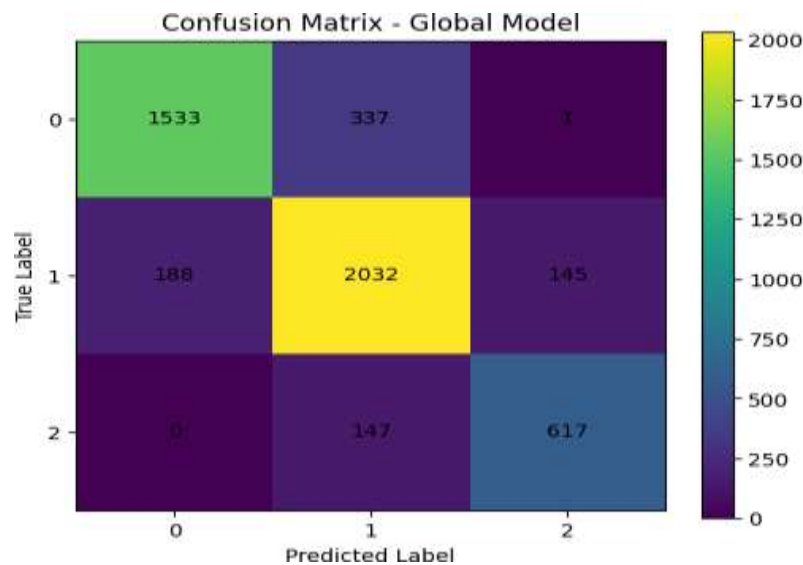
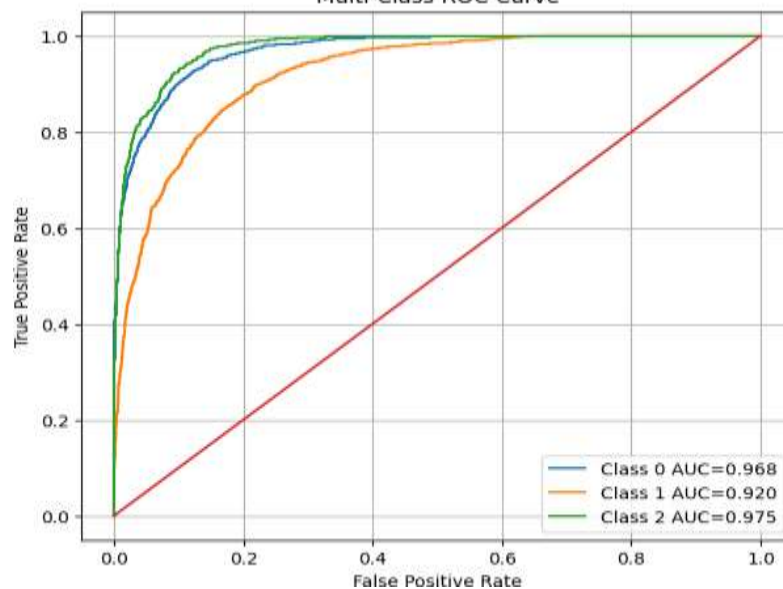


Figure 2: Confusion Matrix of the proposed method.

The performance confusion matrix report in the following image further confirms the classification results, showing that the model correctly classified 1533, 2032, and 617 samples for classes 0, 1, and 2, respectively. The low number of off-diagonal predictions indicates that most of the test samples were correctly classified. The main misclassification was between Class 0 and Class 1, and between Class 1 and Class 2. The classification report shows that the global model achieves strong, relatively balanced predictive performance across all three classes, with class-wise F1-scores of 0.85, 0.83, and 0.81 for Classes 0, 1, and 2, respectively. The overall classification performance on the test set is consistent with a weighted mean of 0.84 for precision, recall and F1-score. The model's overall accuracy is 0.84. The macro-average F1-score of 0.83 indicates that the model is more balanced across the three classes, rather than being skewed toward the majority class. Chart 1 illustrates the ROC curve for the proposed Framework. It achieves AUC values of 0.968, 0.920, and 0.975 for Classes 0, 1, and 2, respectively.

Chart 2: ROC Curve of the Proposed ENSEMBLEFEDSHIELD-PRUDA Framework
Multi-Class ROC Curve



5.1. Ablation study Analysis of ENSEMBLEFEDSHIELD-PRUDA framework:

The ablation study assesses the contribution of systematically omitting individual components, such as adaptive clipping, uncertainty estimation, dynamic attenuation, and adaptive differential privacy. This comparison shows which components contribute most to overall model performance and the privacy-utility trade-off.

Table 3: Ablation Study of proposed framework.

Configuration	Adaptive Clipping	Update Relevance Scoring	MC Drop-out	Dynamic Update At-tenuation	Adap-tive DP	Key Observation
Baseline FL	No	No	No	No	No	Standard federated learning without adaptive privacy mechanisms
Model + Adaptive Clipping	Yes	No	No	No	No	Improved optimization stability through adaptive clipping
+ Update Relevance Scoring	Yes	Yes	No	No	No	Better preservation of informative client updates
+ MC Dropout	Yes	Yes	Yes	No	No	Reliable estimation of prediction uncertainty
+ Dynamic Update Attenuation	Yes	Yes	Yes	Yes	No	Reduced influence of uncertain client updates
Proposed Framework	Yes	Yes	Yes	Yes	Yes	Best balance between model utility and privacy preservation

The above ablation study report explains the importance and trustworthiness of the proposed system. It clearly shows the necessity of the proposed framework, which uses adaptive clipping, relevance scoring, and model reliability using uncertainty estimation with update attenuation, followed by round-based adaptive noise with the best model performance among existing baselines, and it also maintains the best balance between model utility and privacy preservation for securing the model updates during aggregation.

VI. Conclusion

This paper demonstrates a novel ENSEMBLEFEDSHIELD-PRUDA framework helps to pre-serve mental health risk prediction. This method integrates moving average adaptive clipping, update relevance scoring, Monte Carlo dropout-based uncertainty estimation, dynamic update at-tenuation, and round-based adaptive differential privacy into a single federated learning frame-work to balance predictive performance and privacy preservation. Unlike traditional federated learning approaches that rely on static clipping thresholds and fixed differential privacy mechanisms, the proposed adaptive framework dynamically controls client updates based on their relevance and prediction uncertainty before applying adaptive privacy protection. Experimental results on the mental health risk dataset, which contains up to 25000 patient records, show that the proposed framework consistently outperforms traditional baselines while providing strong privacy assurance. The ablation study further validated the contribution of each adaptive component. Overall, the proposed approach offers an effective and practical solution for secure, privacy-preserving collaborative mental healthcare analytics. Future work will extend the approach by adding trust-aware secure aggregation and robust defence mechanisms against adversarial attacks in federated learning environments.

VII. Reference

- [1] A. ElZemity and B. Arief, 'Privacy threats and countermeasures in federated learning for Internet of Things: A systematic review', in 2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics, 2024, pp. 331–338.
- [2] V. Kiruthiga and D. K. L. Priya, 'A privacy-preserving framework for mental health prediction using Federated Learning', Int. J. Res. Appl. Sci. Eng. Technol., vol. 13, no. 7, pp. 1474–1480, July 2025.
- [3] W. Wei, L. Liu, Y. Wu, G. Su, and A. Iyengar, 'Gradient-leakage resilient federated learning', in 2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS), 2021, pp. 797–807.
- [4] P. Ruzafa-Alcázar et al., 'Intrusion detection based on privacy-preserving federated learning for the industrial IoT', IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1145–1154, 2021.
- [5] B. Nagy et al., 'Privacy-preserving federated learning and its application to natural language processing', Knowledge-Based Systems, vol. 268, p. 110475, 2023.
- [6] X. F. Alexander, 'Federated Learning for Privacy-Preserving Smart Healthcare: An Architectural Overview', International Journal of Emerging Trends in Engineering and Technology, vol. 1, no. 1, pp. 1–10, 2025.

- [7] H. Su, C. Xiang, and B. Ramesh, 'Towards confidential Chatbot conversations: A decentral-ised federated learning framework', The Journal of The British Blockchain Association, 2024.
- [8] J. Fu, Z. Chen, and X. Han, 'Adap dp-fl: Differentially private federated learning with adaptive noise', in 2022 IEEE international conference on trust, security and privacy in computing and communications (TrustCom), 2022, pp. 656–663.
- [9] M. A. M. Pranto and N. Al Asad, 'A comprehensive model to monitor mental health based on federated learning and deep learning', in 2021 IEEE International Conference on Signal Pro-cessing, Information, Communication & Systems (SPICSCON), 2021, pp. 18–21.
- [10] Z. Wang, Z. Yang, I. Azimi, and A. M. Rahmani, 'Differential private federated transfer learning for mental health monitoring in everyday settings: A case study on stress detection', in 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2024, pp. 1–5.
- [11] P. Dubey, P. Dubey, and P. N. Bokoro, 'Federated learning for privacy-enhanced mental health prediction with multimodal data integration', Computer Methods in Biomechanics and Bi-omedical Engineering: Imaging & Visualization, vol. 13, no. 1, p. 2509672, 2025.
- [12] S. Chen, J. Yang, G. Wang, Z. Wang, H. Yin, and Y. Feng, 'CLFLDP: Communication-efficient layer clipping federated learning with local differential privacy', Journal of Systems Ar-chitecture, vol. 148, p. 103067, 2024.
- [13] Tamilvendhan R, Sivasankar K, Yuvaraj K, Mrs. S. Lavanya . ' Uncertainty-Aware Federated Learning Framework for Privacy-Preserving Clinical Decision Support ', International Journal Of Novel Research And Development (IJNRD) Volume 11, Issue 3, MARCH 2026 | ISSN: 2456-4184
- [14] B. Pfitzner et al., 'Differentially-private federated learning with non-IID data for surgical risk prediction', in 2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC), 2024, pp. 120–129.
- [15] C. Wang, Y. Zhang, N. Gao, and Q. Luo, 'Differential privacy personalized federated learning based on dynamically sparsified client updates', arXiv preprint arXiv:2503. 09192, 2025.
- [16] ALEXANDER, X. FRANCIS. "Federated Learning for Privacy-Preserving Smart Healthcare: An Architectural Overview." International Journal of Emerging Trends in Engineering and Tech-nology 1, no. 1 (2025): 1-10.
- [17] Stylianos Karagiannis, Christoforos Ntantogian, Emmanouil Magkos, Aggeliki Tsohou Mas-tering data privacy: leveraging K-anonymity for robust health data sharing, International Journal of Information Security, March 2024 DOI:10.1007/s10207-024-00838-8
- [18] Pranto, Md Appel Mahmud, and Nafiz Al Asad. "A comprehensive model to monitor mental health based on federated learning and deep learning." In 2021 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), pp. 18-21. IEEE, 2021.
- [19] V. Uma Rani, M. Sreenivasa Rao "PrivGuard-Sensitivity guided Anonymization based PPDM with Automatic selection of sensitive Attributes" 2021 7th International Conference on Advanced Computing and Communication system (ICACCS).
- [20] Stylianos Karagiannis, Christoforos Ntantogian, Emmanouil Magkos, Aggeliki Tsohou Mas-tering data privacy: leveraging K-anonymity for robust health data sharing, International Journal of Information Security, March 2024 DOI:10.1007/s10207-024-00838-8.