International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(20s) 2026 pp. 1306-1316
https://doi.org/10.70917/ijcisim-2026-5246
Article

# A Comparative Assessment of Adaptive Differential Privacy Preserving Mechanism for Secure Federated Mental Health Classification

**M. Malathi[1], Dr. M. RameshKumar[2]**

[1]Research Scholar, PG & Research Department of Computer Science, Government Arts College, Nandanam, Chennai 600035, Tamil Nadu, India. Email: malathimano2021@gmail.com
[2]Associate Professor and Head, PG & Research Department of Computer Science, Government Arts College, Nandanam, Chennai 600035, Tamil Nadu, India. Email: proframeshkumar@gmail.com

**Abstract:** Mental health applications consist of sensitive and confidential information; maintaining privacy protection is a significant concern. Federated Learning (FL) is an effective privacy-preserving approach that helps to train models without exposing raw data. During processing of model updates, sensitive information may potentially be revealed; therefore, additional security protection measures are necessary. Differential Privacy (DP) is one of the privacy-preserving techniques that inject noise to the sensitive data, and it helps to reduce such risks, but there are still some disadvantages to the adaptive mechanisms. This paper thoroughly evaluated a combination of various adaptive methods that enabled DP with FL. It processes five combinations of adaptive methods, such as gradient-norm-based clipping with time-decay adaptive noise, percentile-based clipping with gradient-norm-based adaptive noise, Moving average adaptive clipping with round-based adaptive noise, layer-wise clipping with layer-wise adaptive noise, and median clipping with epoch-based noise scheduling. Using the DASS-21 mental health patient dataset, it assesses the privacy-utility trade-off through Accuracy, AUC, Precision, Recall, F1 score, Clipping behavior, and Privacy budget analysis. The experimental findings demonstrate that adaptive privacy strategies have a substantial impact on both model performance and privacy preservation; moving-average adaptive clipping and round-based adaptive noise combination achieve a more balanced result than others. These findings provide a strong foundation for developing advanced privacy-aware federated learning frameworks for highly delicate healthcare applications.
**Keywords:** Federated Learning, Differential Privacy, Adaptive Clipping, Adaptive Noise Injection, Privacy-Preserving Machine Learning, Mental Health Analytics

## 1. INTRODUCTION

Federated learning (FL) is a distributed learning paradigm that supports collaborative model training without directly transmitting sensitive data, especially suitable for healthcare applications that need privacy, such as mental health assessment. However, patient raw records remain on local devices; the transfer of model updates through federated optimisation may still leak sensitive information through attacks such as gradient leakage, model inversion, and membership inference. At last, differential privacy is one of the privacy-preserving techniques which helps to protect client updates in federated learning.

A recent research study highlights the advantages of differential privacy largely depend on the clipping technique used to bound model updates and the noise mechanism utilised to preserve privacy. However, the majority of existing approaches handle individual clipping or adaptive noise mechanisms independently, making it challenging to identify which combination of adaptive mechanisms provides the best balance between classification performance and privacy protection. Additionally, distinct optimisation behaviours across communication cycles are displayed by adaptive clipping and adaptive noise mechanisms, leading to variations in model convergence and predictive accuracy.

Based on the challenges, this research article handles a thorough analysis of adaptive methods. A comprehensive comparative evaluation was conducted on five adaptive differential privacy frameworks in a single federated learning environment. To ensure a fair comparison, each framework carries a different combination of clipping and adaptive noise techniques while maintaining the same dataset, communication rounds, client partitions, and aggregation strategy and model setup. The evaluated adaptive methods are gradient-norm clipping with time-decay adaptive noise, percentile-based clipping with gradient-norm adaptive noise, moving-average adaptive clipping with round-based adaptive noise and median clipping with epoch-based noise scheduling. The extensive experiment was conducted on the DASS-21 mental health dataset to analyse each framework in terms of privacy preservation and classification performance. The evaluated results demonstrate that the combination of moving-average adaptive with round-based adaptive noise techniques achieves the best privacy-utility trade-off by generating superior predictive performance while maintaining effective differential privacy protection. These research findings provide a strong adaptive privacy baseline for the development of more advanced privacy-preserving federated learning frameworks.

The main contributions of this research are summarised as follows:

- A thorough comparison of five pairs of adaptive differential privacy mechanisms in a single federated learning framework.
- Evaluate the effectiveness of various adaptive noise scheduling methods on model utility and privacy preservation.
- These techniques are evaluated by the DASS-21 dataset which consists 1,000 patient's data.
- Extensive experimental analysis using metrics such as accuracy, precision, recall, F1-score, AUC, clipping behaviour, and privacy budget measurements.
- To identify the most effective adaptive privacy-preserving strategy to use as a basis for future secure, safe and personalised federated learning frameworks.

## 2. RELATED LITERATURE

Mental health issues are one of the major public health challenges that should affect sensitive information, such as an individual's emotional, psychological, and social well-being [1]. The rapid evolution of machine learning and artificial intelligence has enabled the advancement of intelligent healthcare systems that can process large amounts of sensitive medical and psychological data [2]. For instance, mental health evaluation methods increasingly rely on data-driven models to detect conditions such as anxiety, stress, and depression. However, the centralized acquisition of sensitive patient information raises fundamental privacy concerns and frequently restricts data sharing among institutions.

Federated Learning (FL) is a potential distributed learning approach that allows several clients to collaboratively build a machine learning model without exposing their raw data. By exchanging only model updates with a central server, FL lowers the risks associated with centralized data storage [3]. Even with this advantage, recent research has shown that model updates can still reveal sensitive information through gradient leakage, membership inference, model inversion, and reconstruction assaults [5] [6] [7]. As a result, maintaining security in federated learning environments remains a critical challenge.

Differential privacy (DP) is one of the most widely used solutions for securing federated learning systems. DP enables formal privacy assurance by controlling the influence of single individual data records through noise injection and an update clipping framework. However, inappropriate noise addition can seriously degrade convergence results and model utility, while excessive clipping may eliminate valuable learning information. Therefore, there is still a research problem in finding a practical balance between predictive performance and privacy protection. Recent research studies have explored various adaptive privacy-preserving techniques, such as gradient-based clipping, percentile clipping, adaptive clipping thresholds, layer-wise clipping, and dynamic noise scheduling algorithms. Even if these methods improve the trade-off between privacy and utility, there remains limited comparative study of various

adaptive clipping and adaptive noise techniques under a unified federated learning framework, especially for sensitive mental health applications [10] [11].

In order to address the research gap, this paper offers a thorough comparative investigation of five adaptive differential privacy-enabled federated learning frameworks. The proposed research evaluates Gradient-Norm Based Clipping with Time-Decay Adaptive Noise, Percentile-Based Clipping with Gradient-Aware Noise, Moving-Average Adaptive Clipping with Round-Based Noise, Layer-Wise Clipping with Layer-Wise Noise Injection, and Median Clipping with Epoch-Based Noise Scheduling. These methods are evaluated with the help of the DASS-21 mental health dataset under a consistent federated learning environment. The DASS is an open-source mental health-related dataset that contains 1,000 data samples that are categorized into three severity classes, such as Normal, Mild, and Severe. An 80-20% training and testing split with stratified sampling was employed, where 800 samples were used for model training and 200 samples were used for performance. Logistic regression is a classification model used within a federated learning framework and to emulate decentralized Learning, the training data were dispersed among several simulated clients.

# 3. METHODOLOGY

This research carried out a framework that integrated Federated Learning (FL) with Adaptive Differential Privacy (ADP) to preserve privacy in mental health classification using the DASS-21 dataset. To preserve privacy in the dataset, these workflows make up the system, which retains a common federated learning workflow while using a variety of clipping and noise scheduling techniques. The dataset was pre-processed to improve data quality and ensure reliable model training. The pre-processing pipeline included handling missing values, one-hot encoding of categorical attributes, and feature standardization to eliminate scale variations. Subsequently, the processed dataset was partitioned into training and testing subsets for performance evaluation under the proposed federated learning framework. Logistic regression is a supervised machine learning algorithm used to predict the categorical target variable used for training without exchanging raw data. Upon completion of local model training, adaptive clipping is applied to bound the sensitivity of model updates and prevent excessive parameter variations. Subsequently, an adaptive differential privacy noise mechanism is incorporated into the clipped model updates before transmission to the central server. This process effectively protects sensitive information against potential privacy inference attacks while maintaining an appropriate balance between model utility and privacy preservation. The global model is constructed at the central server by combining the privacy-preserving local updates using Federated Averaging (FedAvg). At last, various privacy metrics like accuracy, precision, recall, F1-score, and AUC curve are used to evaluate the result of the global model. Let’s see how federated learning frameworks work on datasets for processing further adaptive methods. Below, the following local model updates, gradient norm, federated averaging, softmax classification, and final prediction demonstrate how it is evaluated in the following adaptive clipping and noise techniques.

## *3.1 Federated Learning Framework*

Federated learning is a decentralized machine learning technique that trains AI models directly on devices such as smartphones or local servers without exposing raw data. The training dataset is divided among K federated clients, where each client autonomously performs feature standardization and develops a local regression classifier. During each communication round, only the model parameters are transmitted to the central server, ensuring that sensitive patient information is kept locally stored. Now, let's analyse how the following Local Model Update, Gradient Norm Computation, Federated Model Aggregation, Softmax Classification, and Final Prediction work one by one within the Federated Learning framework:

### 3.1.1 Local Model update

Once local training is done, each client calculates the difference between its updated local model and the received global model. The knowledge learned from the client’s private dataset is captured in this local model update,

which is sent to the federated server for global aggregation. Below is the mathematical equation for local model updates.

$$\Delta W_k^{(t)} = W_k^{(t)} - W_g^{(t)}$$

Where $W_k^{(t)}$ represents the local model parameters of the Kth client at communication round t, and $W_g^{(t)}$ represents the appropriate global model parameters.

### 3.1.2 Gradient Norm Computation:

Before transmitting the local updates, the magnitude of each update is quantified using the gradient norm. Because it establishes whether the model update should be clipped to reduce its sensitivity and enhance differential privacy, this measurement is crucial for adaptive clipping systems.

$$||\Delta W_k^{(t)}||_2 = \sqrt{\sum_{i=1}^{d} \left(\Delta W_{k,i}^{(t)}\right)^2}$$

Where d stands for the total number of model parameters that can be trained.

### 3.1.3 Federated Model Aggregation:

The protected model updates are transmitted to the central server once all the client participants have finished local training. The server aggregates these updates using the Federated Averaging (FedAvg) algorithm to generate an improved global model for the next communication round.

$$W_g^{(t+1)} = W_g^{(t)} + \frac{1}{K}\sum_{k=1}^{K} \Delta W_k$$

Where K represents total count of participating federated clients.

### 3.1.4 Softmax Classification:

The aggregated global model predicts the probability of each mental health class using the softmax activation function. This probabilistic formulation enables the classifier to estimate the likelihood of every possible class for a given input sample.

$$P(y = c\backslash x) = \frac{\exp(w_c^T x + b_{c)}}{\sum_{j=1}^{c} \exp(w_j^T x + b_{j)}}$$

Where c denotes the total number of target classes.

### 3.1.5 Final prediction:

The predicted class label is obtained by selecting the class with the highest posterior probability. This final prediction represents the mental health category assigned to the input instance by the federated global model.

$\hat{y} =$ $arg\ \max\ (y = c|x)$

$c$

The following subsections describe the implementation procedure and mathematical formulation of each method used in the comparative analysis.

## *3.2 Method 1: Gradient-Norm Based Clipping with Time-Decay Adaptive Noise:*

Method 1 integrates Gradient-Norm Based Clipping with Time-Decay Adaptive Noise to secure privacy during federated learning. After local model training, the client updates are trimmed according to their gradient norm to reduce sensitivity before transmission. Subsequently, time-decay adaptive noise is integrated into the clipped updates, where the privacy budget steadily grows throughout communication rounds, hence eliminating unneeded disruption in later phases of training. The secured updates are then added using the Federated Averaging (FedAvg) algorithm to create the global model. Lastly, conventional performance criteria are used to assess the classification performance and privacy preservation capacity.

$$\Delta W_k^{clip} = \Delta W_k . min\left(1, \frac{C}{||\Delta W_k||_2}\right)$$

Where $\Delta W_k$ – Local model update, C – Gradient clipping threshold, and $||\Delta W_k||_2$– Euclidean norm of the local model update

### *3.3 Method 2: Percentile-Based Clipping with Gradient-Norm Based Adaptive Noise*

Method 2 uses Percentile-Based Clipping with Gradient-Norm Based Adaptive Noise providing adaptive privacy protection during the federated learning process. Adaptive control over update sensitivity is made possible by the automatic determination of the clipping threshold from the percentile of the model update distribution after local model training. The protected model updates are then sent to the central server for Federated Averaging (FedAvg) after gradient norm-based adaptive noise is integrated in accordance with the gradient norm. Finally, conventional classification and privacy-related performance criteria are used to assess the combined global model.

$Cp = Percen$t(|$\Delta Wk$ |,p)

Cp – Percentile-based clipping threshold ,ΔWk– Local model update, and p – Selected percentile (90th percentile in this study)

### *3.4 Method 3: Moving-Average Adaptive Clipping with Round-Based Adaptive Noise:*

Method 3 improves the privacy-utility trade-off in federated learning by combining Round-Based Adaptive Noise with Moving-Average Adaptive Clipping. Stable and adaptive management of model update sensitivity is made possible by adaptively updating the clipping threshold using the moving average of prior clipping bounds after local model training. The clipped updates are then subjected to round-based adaptive differential privacy noise before being sent to the central server. Then, the global model is built by aggregating the protected updates using Federated Averaging (FedAvg) algorithm. Finally, the performance of the model is assessed using common categorization and privacy-related criteria.

$Ct = \alpha Ct-1 + (1 - \alpha)||\Delta Wk||2$

Where, Ct – Current moving-average clipping bound, Ct−1 – Previous clipping bound, α– Moving-average coefficient, and ‖ΔWk‖2– Euclidean norm of the local model update

### *3.5 Method 4: Layer-Wise Clipping with Layer-Wise Noise*

Method 4 uses Layer-Wise Clipping together with Layer-Wise Noise injection to offers privacy protection during federated learning. Following local model training, the model updates are split into many layers, and each layer is independently clipped according to its corresponding clipping bound to regulate update sensitivity. Each clipped layer is then injected with adaptive layer wise differential privacy noise prior to transmission to the central server. The global model is then built by aggregating the protected layer-wise updates using the Federated Averaging (FedAvg) technique. Standard classification and privacy-related performance criteria are used to assess the final model.

$$\Delta W_l^{clip} = \Delta W_l . \min\left(\frac{C_l}{||\Delta W_l||_2}\right)$$

$\Delta W_l$ – Model update of the lth layer, $C_l$ – Layer-specific clipping bound and $||\Delta W_l||_2$ – Euclidean norm of the lth layer update

### 3.6 **Method 5: Median Clipping with Epoch-Based Noise Scheduling**

Method 5 adopts the Epoch-Based Noise Scheduling with Median Clipping to ensure the privacy protection in federated learning. The median of the absolute model updates is used for estimating the clipping threshold after local model training, which is a reliable way to reduce the influence of extreme parameter fluctuations. Then a fixed privacy budget ((epsilon = 1.0)) is reserved while adding differential privacy noise at each local training period. The Federated Averaging (FedAvg) technique is then used to aggregate the protected model updates in order to form the global model. Finally, performance indicators with respect to privacy and standard classification are used to evaluate the effectiveness of the proposed method.

$$C = median(|\Delta W_k\ |)$$

Where C – Median clipping threshold and $\Delta W_k$ – Local model update

## 4. EXPERIMENTAL SETUP

### *4.1 Dataset Description*

The proposed experimental evaluations are performed in the DASS-21 Mental Health Dataset, which consists of 1000 participants. The dataset contains demographic information and psychological evaluation responses that have behavioural attributes such as Depression, Anxiety, and Stress scale. In this mental health category, there are classifications such as "Normal," "Mild," and "Severe," representing increasing levels of psychological distress. For reliable model performance, the dataset was partitioned into 80% training and 20% testing sets. Before model training, one-hot encoding was performed on categorical attributes, while numerical features were standardized to negate scale variations. The training data were subsequently distributed among multiple federated clients, where local model training was carried out independently without sharing raw patient-sensitive data, thereby protecting data privacy throughout the federated learning process.

### *4.2 Environmental Setup*

The proposed research work was carried out with the help of Python 3.11 using Pandas, NumPy, Scikit-learn, and Matplotlib libraries. For data loading and pre-processing, Pandas was used; NumPy was used for numerical computations, for model training, and for performance evaluation. Scikit-learn was used, and Matplotlib for visualization of the experimental results. All five adaptive combinations of methods were processed in an identical software environment to ensure a fair performance comparison.

### *4.3 Data Pre-processing*

Data pre-processing is an essential feature for cleaning the dataset to ensure consistency and enhance model performance. In this process numerical features were standardized using a standard scaler, one-hot encoding was performed on categorical attributes, and missing values were replaced with zero. Minimizing the data leakage, a separate global scaler was used for the final assessment after client-specific scaling was implemented during federated learning.

### *4.4 Federated Learning Configuration*

This research framework uses a client-server federated learning architecture, where the training dataset is evenly partitioned among three federated clients. Each client can independently train a local logistic regression model using its own dataset while differential privacy is applied to secure model updates before transmission. The secured updates are aggregated at the central server using the Federated Averaging (FedAvg) algorithm over ten communication

rounds. For a fair and unbiased comparison, communication rounds, training parameters, identical client partitions, and aggregation settings were maintained across all adaptive differential privacy methods.

### *4.5 Model Configuration*

A logistic regression (LR) classifier was used as the baseline model for its computational effectiveness in federated learning. During local training, the classifier was configured with a maximum of 100 optimization iterations, balanced class weighting, and warm-start activation to enhance convergence and maintain stable model updates throughout subsequent communication rounds.

### *4.6 Adaptive Differential Privacy*

Five adaptive differential privacy mechanisms were assessed and evaluated within the same federated learning framework. Each method evaluates a distinct combination of clipping and adaptive noise techniques to investigate their impact on privacy preservation and model performance. To ensure fair comparative analysis, all experimental conditions, including the federated architecture, dataset partition, communication rounds, model configuration, and evaluation metrics, remained at the same level. The below table highlights the corresponding adaptive clipping and adaptive noise strategies that were employed in this research.

**Table 1:** *List of Adaptive Methods performed in this research.*

| **Methods** | **Clipping Strategy** | **Noise Strategy** |
|---|---|---|
| Method 1 | Gradient-Norm based Clipping | Time Decay Adaptive Noise |
| Method 2 | Percentile-Based Clipping | Gradient-Norm based Adaptive Noise |
| Method 3 | Moving-average Adaptive Clipping | Round-Based Adaptive - Noise |
| Method 4 | Layer-wise clipping | Layer wise Noise |
| Method 5 | Median clipping | Epoch based Noise Scheduling |

### *4.7 Performance Evaluation:*

The performance analysis of these techniques is measured using standard classification metrics, such as accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC), and confusion matrix. In addition to predictive performance, privacy-related metrics, including clipping threshold, gradient norm moving clipping bound, layer-wise clipping bound, and privacy budget (ε), were recorded according to the corresponding differential privacy mechanisms. These evaluation analyses assess the comprehensive assessment of the effectiveness of each method in achieving an optimal trade-off between classification performance and privacy preservation under the federated learning framework.

## 5. RESULT

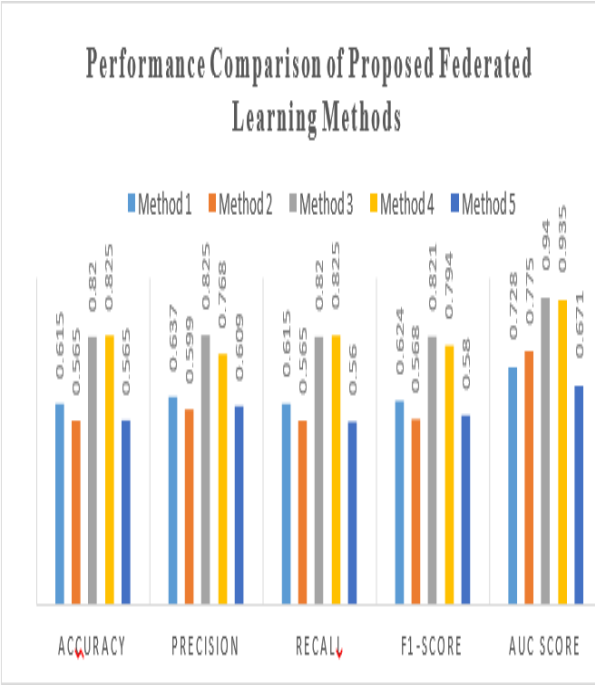
The evaluation comparison is majorly focused on finding the best adaptive combination, which is best for performance as well as the privacy-utility trade-off. Five combinations of adaptive clipping and adaptive noise mechanisms are evaluated in the DASS-21 dataset in the federated learning framework. The assessed methods are gradient-norm-based clipping with time-decay adaptive noise, percentile-based clipping with gradient-norm adaptive noise, moving-average adaptive clipping with round-based adaptive noise, layer-wise clipping with layer-wise adaptive noise, and median clipping with epoch-based noise scheduling. The overall metrics performance was

analysed using accuracy, precision, F1-score, Recall, AUC score, Confusion Matrix and Privacy related measures, which include privacy budget and clipping thresholds. The below table shows the performance evaluation of five combinations of adaptive differential privacy clipping with noise evaluated in a federated learning framework.

Table 2: Performance Evaluation of Adaptive Differential Privacy with FL.

| **Methods** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **AUC Score** |
|---|---|---|---|---|---|
| Method 1 | 0.615 | 0.637 | 0.615 | 0.624 | 0.728 |
| Method 2 | 0.565 | 0.599 | 0.565 | 0.568 | 0.775 |
| Method 3 | 0.820 | 0.825 | 0.820 | 0.821 | 0.94 |
| Method 4 | 0.825 | 0.768 | 0.825 | 0.794 | 0.935 |
| Method 5 | 0.565 | 0.609 | 0.56 | 0.58 | 0.671 |

The performance metrics scored from the above result table (accuracy, precision, recall, F1-score, and AUC score) were plotted to clearly demonstrate the impact of each adaptive noise and clipping technique. The below chart visualization depicts the understanding of the trade-off between privacy strength and model performance.




Figures 1–5 represent the execution reports of the five adaptive differential privacy methods (gradient-norm-based clipping with time-decay adaptive noise, percentile-based clipping with gradient-norm adaptive noise, moving-average adaptive clipping with round-based adaptive noise, layer-wise clipping with layer-wise adaptive noise, and median clipping with epoch-based noise) implemented in the proposed federated

```
===== METHOD 1 RESULTS =====
Accuracy  : 0.6150
Precision : 0.6370
Recall    : 0.6150
F1 Score  : 0.6248
AUC Score : 0.7281
Avg Clip Norm : 9.7374
Final Epsilon : 1.0000

Confusion Matrix:
 [[43 15 21]
 [10  4  1]
 [27  3 76]]

Classification Report:
              precision    recall  f1-score   support

        Mild       0.54      0.54      0.54        79
      Normal       0.18      0.27      0.22        15
      Severe       0.78      0.72      0.75       106

    accuracy                           0.61       200
   macro avg       0.50      0.51      0.50       200
weighted avg       0.64      0.61      0.62       200
```

**Fig 1: Analysis report of Method 1.**

```
===== METHOD 2 RESULTS =====
Accuracy  : 0.5650
Precision : 0.5996
Recall    : 0.5650
F1 Score  : 0.5684
AUC Score : 0.7750
Avg Percentile Threshold : 2.0740
Avg Gradient Norm : 2.0740

Confusion Matrix:
 [[38 19 22]
 [ 0 15  0]
 [46  0 60]]

Classification Report:
              precision    recall  f1-score   support

        Mild       0.45      0.48      0.47        79
      Normal       0.44      1.00      0.61        15
      Severe       0.73      0.57      0.64       106

    accuracy                           0.56       200
   macro avg       0.54      0.68      0.57       200
weighted avg       0.60      0.56      0.57       200
```

**Fig 2: Analysis report of Method 2**

```
===== METHOD 3 RESULTS =====
Accuracy  : 0.8200
Precision : 0.8256
Recall    : 0.8200
F1 Score  : 0.8211
AUC Score : 0.9431
Avg Moving Clip Bound : 6.0925
Final Epsilon : 1.5000

Confusion Matrix:
 [[65  1 13]
 [ 4 11  0]
 [18  0 88]]

Classification Report:
              precision    recall  f1-score   support

        Mild       0.75      0.82      0.78        79
      Normal       0.92      0.73      0.81        15
      Severe       0.87      0.83      0.85       106

    accuracy                           0.82       200
   macro avg       0.85      0.80      0.82       200
weighted avg       0.83      0.82      0.82       200
```

**Fig 3: Analysis report of Method 3**

```
===== METHOD 4 RESULTS =====
Accuracy  : 0.8250
Precision : 0.7680
Recall    : 0.8250
F1 Score  : 0.7946
AUC Score : 0.9359
Avg Layer Clip Bound : 0.1328

Confusion Matrix:
 [[68  0 11]
 [15  0  0]
 [ 9  0 97]]

Classification Report:
              precision    recall  f1-score   support

        Mild       0.74      0.86      0.80        79
      Normal       0.00      0.00      0.00        15
      Severe       0.90      0.92      0.91       106

    accuracy                           0.82       200
   macro avg       0.55      0.59      0.57       200
weighted avg       0.77      0.82      0.79       200
```

**Fig 4: Analysis report of Method 4**

```
===== METHOD 5 RESULTS =====
Accuracy  : 0.5650
Precision : 0.6098
Recall    : 0.5650
F1 Score  : 0.5824
AUC Score : 0.6719
Avg Median Clip Bound : 1.4332
Privacy Budget ε : 1.0000

Confusion Matrix:
 [[41 14 24]
 [ 7  6  2]
 [27 13 66]]

Classification Report:
              precision    recall  f1-score   support

        Mild       0.55      0.52      0.53        79
      Normal       0.18      0.40      0.25        15
      Severe       0.72      0.62      0.67       106

    accuracy                           0.56       200
   macro avg       0.48      0.51      0.48       200
weighted avg       0.61      0.56      0.58       200
```

**Fig 5: Analysis report of Method 5.**

Method 1 provides the best baseline by integrating gradient clipping with time decay noise scheduling. Method 2 enhanced the baseline by using gradient-aware noise creation and adaptive clipping. To enhance further the privacy-utility trade-off by using moving-average clipping together with round-based adaptive noise, it produced the most balanced classification performance and model convergence. Methods 4 and 5 also produced effective privacy preservation through layer-wise and median-based strategies, but their predictive performance is less active than Method 3. From the overall result, Method 3 indicates that it improves privacy-preserving federated learning for mental health prediction compared to other combinations of methods. This leads to better convergence of the model and preserves strong privacy, resulting the highest accuracy, precision, recall, F1 score and AUC. While other methods such as 1, 2, 4, and 5 also effectively maintain privacy, their noise and clipping strategies caused comparatively higher perturbation, which reduced predictive performance. Overall the comparative result of Method 3 demonstrates that adaptive clipping combined with adaptive noise scheduling significantly enhances federated learning effectiveness for mental health classification.

## 6. CONCLUSION AND FUTURE WORK

This research contribution provides a detailed evaluation of five adaptive differential privacy mechanisms integrated with federated learning for privacy-preserving mental health classification on the DASS-21 dataset. The

proposed framework achieved better classification performance and effectively protected sensitive client information by using adaptive clipping and adaptive noise injection. When compared to conventional differential privacy methods, experimental outcomes demonstrate that moving average clipping with round based noise adaptive privacy mechanisms provide a superior balance between maintaining privacy and predictive performance. Among all the comparisons, Moving-Average Adaptive Clipping with Round-Based Adaptive Noise (Method 3), which also achieves the best overall privacy-utility trade-off, shows superior model convergence and robust classification performance. Future research will focus on expanding and extending the framework to federated learning models based on deep learning, providing customized client-specific adaptive privacy metrics, and validating the system on wider multi-institutional healthcare datasets. Future direction may look into communication-efficient optimization procedures and secure aggregation strategies to improve scalability, privacy protection, and practical deployment in remote healthcare applications.

## REFERENCES


1. W.H. Organization, *World mental health report: transforming mental health for all. Executive summary*. World Health Organization, 2022.
2. T. M. Mitchell and T. M. Mitchell, *Machine learning*, vol. 1. McGraw-hill New York, 1997.
3. 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.
4. J. Konečn\`y, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, 'Federated learning: Strategies for improving communication efficiency', *arXiv preprint arXiv:1610. 05492*, 2016..
5. L. Zhu, Z. Liu, and S. Han, 'Deep leakage from gradients', *Advances in neural information processing systems*, vol. 32, 2019.
6. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, 'Membership inference attacks against machine learning models', in *2017 IEEE symposium on security and privacy (SP)*, 2017, pp. 3–18.
7. M. Fredrikson, S. Jha, and T. Ristenpart, 'Model inversion attacks that exploit confidence information and basic countermeasures', in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.
8. C. Dwork, 'Differential privacy', in *International colloquium on automata, languages, and programming*, 2006, pp. 1–12.
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 Processing, Information, Communication & Systems (SPICSCON)*, 2021, pp. 18–21.
10. M. H. Fares and A. M. S. E. Saad, 'Towards privacy-preserving medical imaging: Federated learning with diffrential privacy and secure aggregation using a modified resnet architecture', *arXiv preprint arXiv:2412. 00687*, 2024.
11. [11]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.
12. V. U. Rani and M. S. Rao, 'Privguard: Sensitivity guided anonymization based ppdm with automatic selection of sensitive attributes', in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2021, vol. 1, pp. 1832–1837.
13. S. Karagiannis, C. Ntantogian, E. Magkos, A. Tsohou, and L. L. Ribeiro, 'Mastering data privacy: leveraging K-anonymity for robust health data sharing: Mastering data privacy: leveraging K-anonymity..', *International Journal of Information Security*, vol. 23, no. 3, pp. 2189–2201, 2024.
14. 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 Processing, Information, Communication & Systems (SPICSCON)*, 2021, pp. 18–21.
15. 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.
16. Stylianos Karagiannis, Christoforos Ntantogian, Emmanouil Magkos,Aggeliki TsohouMastering data privacy: leveraging K-anonymity for robust health data sharing, International Journal of Information Security, March 2024 DOI:10.1007/s10207-024-00838-8
17. [17] 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.