


Privacy Preserving Machine Learning With Federated Personalized Learning in Artificially Generated Environment

MD. TANZIB HOSAIN ^{1,2}, MUSHFIQUR RAHMAN ABIR¹, MD. YEASIN RAHAT¹,
M. F. MRIDHA ^{1,2} (Senior Member, IEEE), AND SADDAM HOSSAIN MUKTA ³

¹Department of Computer Science & Engineering, American International University-Bangladesh, Dhaka 1229, Bangladesh

²Advanced Machine Intelligence Research Lab, Dhaka 1207, Bangladesh

³LUT School of Engineering Sciences, LUT University, 53850 Lappeenranta, Finland

CORRESPONDING AUTHOR: SADDAM HOSSAIN MUKTA (e-mail: saddam.mukta@lut.fi).

ABSTRACT The widespread adoption of Privacy Preserving Machine Learning (PPML) with Federated Personalized Learning (FPL) has been driven by significant advances in intelligent systems research. This progress has raised concerns about data privacy in the artificially generated environment, leading to growing awareness of the need for privacy-preserving solutions. There has been a seismic shift in interest towards Federated Personalized Learning (FPL), which is the leading paradigm for training Machine Learning (ML) models on decentralized data silos while maintaining data privacy. This research article presents a comprehensive analysis of a cutting-edge approach to personalize ML models while preserving privacy, achieved through the innovative framework of Privacy Preserving Machine Learning with Federated Personalized Learning (PPMLFPL). Regarding the increasing concerns about data privacy in virtual environments, this study evaluated the effectiveness of PPMLFPL in addressing the critical balance between personalized model refinement and maintaining the confidentiality of individual user data. According to our results based on various effectiveness metrics, the use of the Adaptive Personalized Cross-Silo Federated Learning with Homomorphic Encryption (APPLE+HE) algorithm for privacy-preserving machine learning tasks in federated personalized learning settings within the artificially generated environment is strongly recommended, obtaining an accuracy of 99.34%.

INDEX TERMS Extended reality, federated personalized learning, privacy, privacy preserving machine learning, security.

I. INTRODUCTION

Traditional Machine Learning (ML) models are often centralized, in which all data are collected and stored in a single location for training. This centralization exposes users' personal information to potential breaches and misuse, leading to privacy infringement [1]. Privacy concerns in ML have been further exacerbated by the development of Deep Learning (DL) models, which require even more data to achieve state-of-the-art performance. Within the artificially generated environment, engaging users in virtual environments often share vast amounts of personal data, including behavioral patterns, preferences, and even biometric information, which can be collected, stored, and potentially exploited by platform

operators, advertisers, and malicious actors. Moreover, the persistent nature of virtual identities and activities in the artificially generated environment amplifies the stakes for privacy, necessitating robust frameworks and technologies for ensuring data protection, anonymity, and user control over personal information [2].

On the other hand, the exponential growth of data and advancements in ML have revolutionized various domains, offering personalized services and recommendations tailored to individual users. However, this progress has raised significant privacy concerns regarding the collection, storage, and processing of sensitive user data [3]. The conventional approach of centralizing data for training machine learning models

exposes users' personal information to potential breaches and misuse, causing valid apprehensions regarding data privacy [4]. Consequently, there is an increasing requirement for Privacy Preserving Machine Learning (PPML) techniques that ensure personalized services, while safeguarding individual data privacy.

Several privacy-preserving techniques have been proposed to address privacy concerns, including Differential Privacy (DP), Homomorphic Encryption (HE), and Secure Multi-Party Computation (SMPC) [5]. These methods aim to protect user data during model training and inference; however, they often face challenges in maintaining model accuracy and scalability. A promising solution to this dilemma is the concept of Federated Learning (FL), which allows decentralized model training across multiple devices while keeping raw data local and secure [1]. FL enables multiple iterations of model updates without exposing sensitive data, thereby ensuring a robust privacy guarantee. However, standard FL models face challenges such as handling heterogeneous data, privacy concerns from aggregating model updates, suboptimal model performance for individual users, communication inefficiency, and scalability issues [6]. Personalized Federated Learning (pFL) addresses these issues by creating models tailored to individual users' data, thus enhancing model performance and user satisfaction while reducing the need for extensive data sharing [7]. In addition, pFL can employ decentralized methods to further reduce privacy risks and communication overhead, making it a compelling choice for scenarios with significant data heterogeneity and privacy concerns [8].

The need to strike a delicate balance between delivering personalized experiences and safeguarding individual privacy has led to the emergence of innovative framework. For this case, APPLE+HE will be a paradigm shift in the realm of privacy-preserving model personalization while ensuring that user data remain localized and secure. As far as our findings, no previous works had done with the integration of PPML and FPL on the data of virtual environments. We think our contribution provides potential impacts on providing a new PPMLFPL benchmarks for artificially generated environment.

The overall contributions of this study are-

- This study provides a thorough experiment with personalized machine learning models while preserving privacy.
- We assess the effectiveness of the PPMLFPL framework in maintaining the balance between personalized model refinement and user data confidentiality in virtual environments.
- According to the empirical results, this study strongly recommends using the APPLE+HE algorithm for privacy-preserving machine learning tasks.
- Also, this study explains why APPLE+HE performs better, emphasizing the combination of personalized federated learning capabilities of the APPLE algorithm and the robust privacy guarantee provided by HE.
- At last, this study outlines future research works, emphasizing the need to explore the scalability of APPLE+HE

in larger and more diverse virtual environments, its resilience against sophisticated privacy attacks, real-time adaptation to dynamic data distributions.

The rest of the article is as follows: Section II reviews the existing literature highlighting the necessities of this study. Then, Section III-A highlights the key differences between federated and personalized federated learning. After that, Section IV discusses the experimental methodology. Later, Section V provides the in detail performance analysis of the experiment. Finally, Section VI briefly discusses the findings and the directs potential future works and thus VI concludes the article.

II. ADVANCEMENTS OF PRIVACY PRESERVING MACHINE LEARNING AND FEDERATED LEARNING

Recent advancements in PPML have led to the development of various techniques for enhancing data security and privacy. These techniques include DP [9], SMPC [10], HE [11], secure enclave technologies [12], privacy preserving deep learning [2], privacy preserving data sharing [13], and privacy preserving data publishing [14]. Each of these techniques encompasses specific methodologies and tools tailored to protect sensitive information. For instance, under DP, noise addition mechanisms such as Laplace and Gaussian Noise, are highlighted, as well as Secure Aggregation (SA) [5] techniques employing SMPC and HE. Additionally, emerging technologies in this field feature HE and secure enclave technologies such as Intel SGX [15], AMD SEV [16], and ARM TrustZone [17], which offer robust environments for secure data processing. These advancements represent significant strides in ensuring privacy and security in machine learning applications.

Personalized FL advances involve FedMTL [18] for multi-task learning, FedBN [19] for handling non-Identically and Independently Distribution (non-IID) features via local batch normalization, and meta-learning-based pFL algorithms such as Per-FedAvg [7] for personalized federated learning with theoretical guarantees. Regularization-based pFL, such as pFedMe [20] and Ditto [21], utilize moreau envelopes and fairness principles for improved personalization. Personalized-aggregation-based pFL approaches like APFL [22], FedFomo [23], FedAMP [24], FedPHP [25], APPLE [26], and FedALA [27] integrate adaptive local aggregation, first-order model optimization, and inherited private models for personalized federated learning. Model-splitting-based pFL techniques such as FedPer [28], LG-FedAvg [29], FedRep [30], FedRoD [31], FedBABU [32], and FedGC [33] exploit shared representations and gradient correction for enhanced personalized federated image classification and face recognition. Knowledge-distillation-based pFL methods, such as FedDistill [34], FML [35], FedKD [6], and FedProto [36] leverage knowledge distillation and mutual learning across clients to improve personalized federated learning. FedPCL [37] and FedPAC [8] introduce contrastive learning and feature alignment with classifier collaboration for further advancements in personalized federated learning. These

advancements collectively address the challenges of privacy, communication, and personalization in federated learning scenarios.

III. FEDERATED LEARNING VS PERSONALIZED FEDERATED LEARNING

A. FEDERATED LEARNING

FL aims to train a global model using decentralized data sources without explicitly exchanging raw data between devices and a central server. Mathematically, the objective of FL is to minimize the global loss $\mathcal{L}_{\text{global}}$, defined as:

$$\min_{\theta} \mathcal{L}_{\text{global}}(\theta) = \sum_{k=1}^K \frac{n_k}{N} \mathcal{L}_k(\theta), \quad (1)$$

where θ represents the global model parameters, $\mathcal{L}_k(\theta)$ denotes the local loss function of each participating device k with n_k representing the number of samples available at device k , and $N = \sum_{k=1}^K n_k$ is the total number of samples across all devices.

B. FEDERATED PERSONALIZED LEARNING

FPL extends FL by allowing customization of the global model to better fit local data characteristics while maintaining privacy. Mathematically, FPL introduces personalization parameters ϕ_k for each device k , leading to a personalized loss function $\mathcal{L}_{\text{personal}}(\theta, \phi_k)$:

$$\min_{\theta, \phi_k} \mathcal{L}_{\text{personal}}(\theta, \phi_k) = \sum_{k=1}^K \frac{n_k}{N} \mathcal{L}_k(\theta, \phi_k), \quad (2)$$

where $\mathcal{L}_k(\theta, \phi_k)$ represents the local loss function for device k that depends on both the global model parameters θ and the personalized parameters ϕ_k . The objective is to optimize $\mathcal{L}_{\text{personal}}(\theta, \phi_k)$ jointly across all devices, enabling customization of the model while preserving the privacy of individual data.

C. KEY DIFFERENCES

1) GLOBAL VS. PERSONALIZED PARAMETERS

FL optimizes a single set of global parameters θ across all devices, aiming to minimize a global loss function averaged over all data sources. In contrast, FPL introduces personalized parameters ϕ_k for each device k , allowing for customization of the model to better fit local data characteristics.

2) OBJECTIVE FUNCTIONS

FL minimizes the global loss function $\mathcal{L}_{\text{global}}(\theta)$, which is a weighted average of local losses across all devices. FPL optimizes the personalized loss function $\mathcal{L}_{\text{personal}}(\theta, \phi_k)$, which includes both the global model parameters θ and personalized parameters ϕ_k specific to each device.

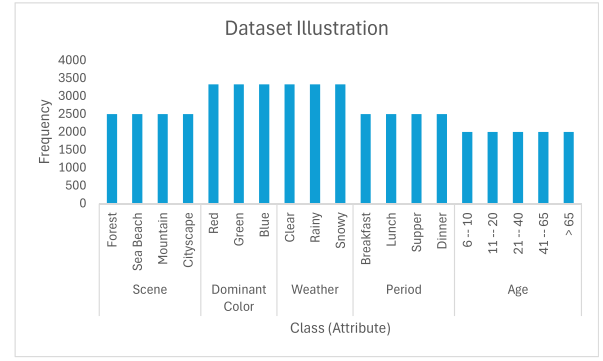


FIGURE 1. Classes with their attributes and respective frequency of the dataset.

3) PRIVACY AND CUSTOMIZATION

FL focuses on aggregating knowledge from distributed data sources while maintaining privacy by not sharing raw data. FPL enhances privacy by allowing each device to optimize personalized parameters while contributing to the global model, thereby balancing model customization with privacy preservation.

IV. METHODOLOGY

A. DATASET

To obtain control over heterogeneous data availability in virtual environments, we created a custom dataset by leveraging Generative Artificial Intelligence (GAI) techniques. Our dataset encompasses five primary classes with 19 distinct attributes aimed at detecting and extracting information from environmental images. After getting an input image, the output attributes include scene type, dominant color, weather conditions, time of day, and the person's age if anyone exists in that image. Information about the building in the dataset is shown in Fig. 1. For image generation, we primarily used Stable Diffusion and DALL-E.

We divided the image generation and image data into five different classes, matching out 19 attributes. Each class comprises 10000 images totaling 50000 images for the entire dataset. This substantial dataset size allowed for sufficient variation and diversity within each attribute category, thereby enhancing the generalizability and effectiveness of our models.

The generation prompt plays a crucial role in image generation. We have cautiously engineered prompts to obtain the right image outputs. For example, “A rural countryside scene with rolling hills, farmland, and wind turbines generating clean energy”. We attempted to make the environmental image as detailed as possible by using specific prompts.

B. EXPERIMENTAL SETUP

To ensure the integrity of our evaluation process in decentralized settings, we randomly partitioned the dataset into

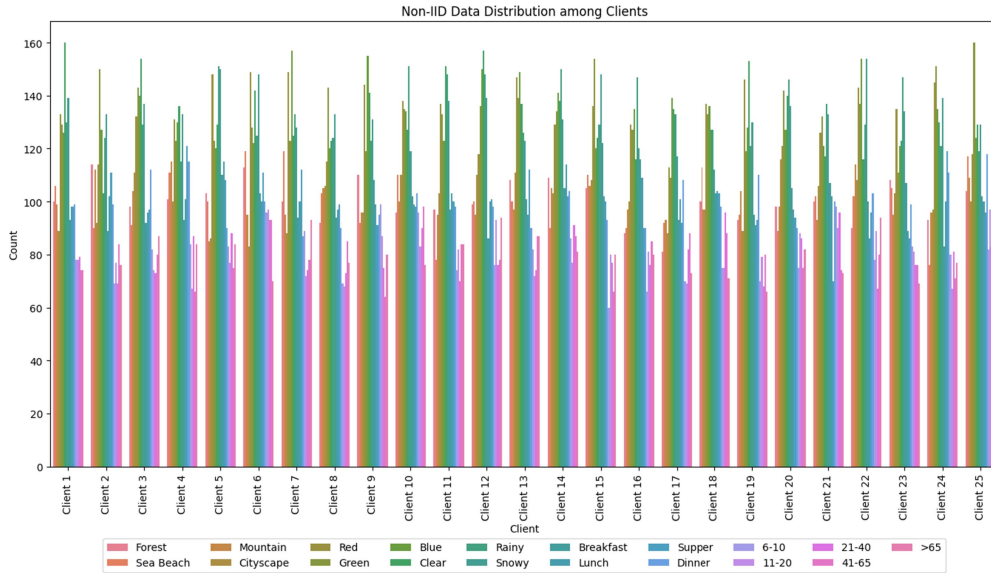


FIGURE 2. Frequency of the non-IID distribution of the dataset across clients. It indicate the number of samples per client splitted in non-IID setting.

training and testing sets, maintaining a non-IID distribution of images across all classes. To conduct performance tests for algorithmic decision making in a distributed setting, we deployed 25 clients across our computational infrastructure, with one machine acting as the server. The details on the demographic diversity of the data across clients are as follows:

- **Age Representation:** The dataset includes a broad age range from children (6–10 years) to elderly individuals (>65 years). This ensures that the model can be trained and evaluated on data representing people from different age groups, which is essential for fairness.
- **Environmental and Cultural Context:** The diversity in scenes (forest, beach, mountain, cityscape) and weather conditions (clear, rainy, snowy) provides a mix of rural, urban, and natural settings. This variety helps the model perform well in diverse cultural and environmental contexts.
- **Temporal Representation:** The inclusion of different periods (breakfast, lunch, supper, dinner) represents various social and daily activities, helping the model generalize across different times of day and social interactions.

The data distribution is shown in Fig. 2. Furthermore, we conducted our experiment using four virtual machines connected via Ethernet cables. Each machine runs on the x86_64 architecture with Arch Linux 2023.04.01 and kernel version 6.2.2-arch1-1. Machines A and B are equipped with Intel i9-9900K CPUs and 31.75 GiB of memory, while machines C and D feature Intel i9-10900K CPUs with 31.76 GiB of memory. All machines operate on the Linux operating system, enabling efficient communication through direct connections between the virtual machines. Following are the implementation details of key components in the experimental setting:

1) FEDERATED PERSONALIZED LEARNING WITH DIFFERENTIAL PRIVACY

Differential privacy is a rigorous mathematical framework designed to provide strong privacy guarantees when performing data analysis [4]. It ensures that the inclusion or exclusion of a single individual's data does not significantly affect the output of a computation, thereby protecting individual privacy [9]. Following are the detailed technical parameters and configurations used in differential privacy:

• Privacy Loss Parameter (ϵ)

- **Definition:** ϵ (epsilon) is the privacy loss parameter that quantifies the privacy guarantee. A smaller ϵ value indicates stronger privacy.
- **Interpretation:** If ϵ is close to zero, the output of the algorithm is almost independent of any single individual's data.
- **Configuration:** Typically, ϵ ranges from 0.01 to 10. Lower values like 0.01 or 0.1 provide strong privacy but might reduce the utility of the data, while higher values offer better utility but weaker privacy.

• Sensitivity (Δf)

- **Definition:** Sensitivity measures how much a single individual's data can change the output of a function. Formally, for a function f , the sensitivity Δf is defined as:

$$\Delta f = \max_{D, D'} \|f(D) - f(D')\| \quad (3)$$

where D and D' are datasets differing by one element.

- **Interpretation:** Higher sensitivity implies greater potential for a single data point to affect the result, necessitating more noise to maintain privacy.
- **Configuration:** Sensitivity depends on the function and the data. For example, for counting queries, the sensitivity is typically 1.

• Noise Distribution

- **Laplace Distribution:** Commonly used due to its simple implementation. The noise added follows a Laplace distribution with mean 0 and scale $\frac{\Delta f}{\epsilon}$:

$$\text{Lap}\left(\frac{\Delta f}{\epsilon}\right) \quad (4)$$

- **Gaussian Distribution:** Used for (ϵ, δ) -differential privacy, where δ is a small probability of failure. The noise added follows a Gaussian distribution with mean 0 and standard deviation σ :

$$\sigma = \frac{\Delta f \sqrt{2 \ln(1.25/\delta)}}{\epsilon} \quad (5)$$

- **Interpretation:** The choice of noise distribution affects the trade-off between privacy and accuracy. Gaussian noise is often used when stronger privacy guarantees are required with a small probability of failure.

• Composition

- **Sequential Composition:** When multiple differentially private algorithms are applied sequentially on the same dataset, the overall privacy loss is the sum of the individual privacy losses:

$$\epsilon_{\text{total}} = \sum_i \epsilon_i \quad (6)$$

- **Parallel Composition:** When differentially private algorithms are applied to disjoint subsets of the dataset, the overall privacy loss is determined by the maximum ϵ among the algorithms.
- **Advanced Composition:** Provides tighter bounds on the privacy loss when multiple algorithms are applied. For (ϵ, δ) -differential privacy, it states that for k algorithms with privacy parameters ϵ and δ , the total privacy guarantee is:

$$(\epsilon \sqrt{2k \ln(1/\delta')} + k\epsilon(e^\epsilon - 1), k\delta + \delta') \quad (7)$$

• Privacy Budget

- **Definition:** The privacy budget is the total allowable privacy loss (ϵ) for a data analysis task or over the lifetime of a dataset.
- **Interpretation:** Managing the privacy budget is crucial for ensuring privacy guarantees are not exceeded. Each query or analysis consumes a portion of the privacy budget.
- **Configuration:** The privacy budget should be carefully allocated to balance between the number of queries and the privacy guarantee.

• Privacy Amplification

- **Subsampling:** When a random sample of data is used, the privacy guarantee can be amplified. If an algorithm is (ϵ, δ) -differentially private, running it on a random subsample of the data provides better privacy guarantees.

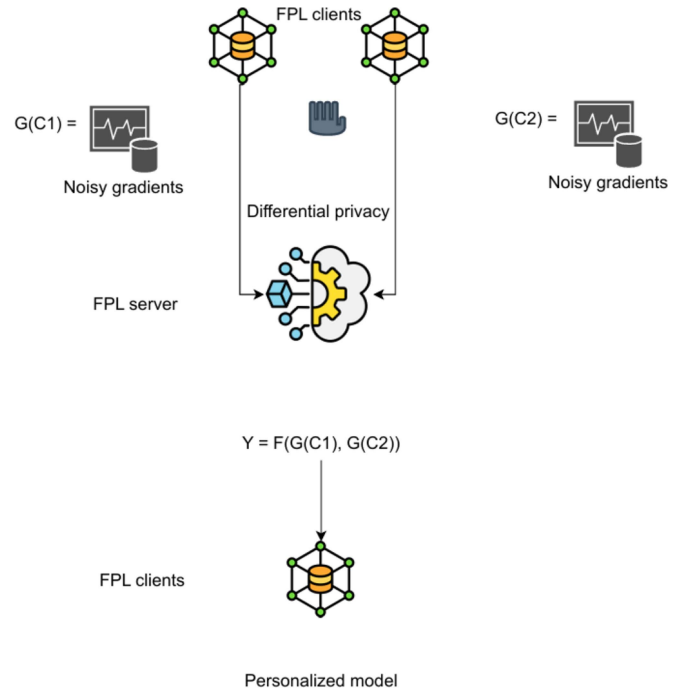


FIGURE 3. Federated personalized learning with Differential Privacy algorithm.

- **Shuffling:** Randomly shuffling the data before applying a local differentially private mechanism can amplify privacy guarantees.

• Post-Processing Invariance

- **Definition:** Any data-independent transformation applied to the output of a differentially private mechanism does not degrade the privacy guarantee. If a mechanism is ϵ -differentially private, any function applied to its output remains ϵ -differentially private.

Fig. 3 illustrates the working procedure of Federated Differential Privacy in detail.

2) FEDERATED PERSONALIZED LEARNING WITH HOMOMORPHIC ENCRYPTION

Homomorphic encryption is a form of encryption that allows computations to be carried out on ciphertext, generating an encrypted result that, when decrypted, matches the result of operations performed on the plaintext [11]. This enables privacy-preserving computation on encrypted data. Following are the detailed technical parameters and configurations used in homomorphic encryption:

• Key Parameters

– Security Parameter (λ)

- * **Definition:** Determines the level of security of the encryption scheme. A larger security parameter implies higher security.
- * **Configuration:** Typical values are 128, 192, or 256 bits.

- * **Interpretation:** Higher values provide better security but require more computational resources.

– Plaintext Space

- * **Definition:** The set of possible plaintext values that can be encrypted.
- * **Configuration:** Often defined as integers modulo some number N , where N is a product of primes.
- * **Interpretation:** The size of the plaintext space affects the types of computations that can be performed.

– Ciphertext Space

- * **Definition:** The set of possible ciphertext values.
- * **Configuration:** Typically larger than the plaintext space to allow for homomorphic operations.
- * **Interpretation:** The ciphertext space must be large enough to prevent overflow during computations.

• Encryption Parameters

– Public Key (pk)

- * **Definition:** Used to encrypt plaintexts.
- * **Configuration:** Includes parameters like large primes and group generators, depending on the encryption scheme.
- * **Interpretation:** Must be distributed securely to users who need to perform encryption.

– Private Key (sk)

- * **Definition:** Used to decrypt ciphertexts.
- * **Configuration:** Includes secret values that correspond to the public key parameters.
- * **Interpretation:** Must be kept secure by the data owner.

• Homomorphic Operations

– Addition (\oplus):

- * **Definition:** The operation that adds two encrypted values.
- * **Configuration:** Depends on the specific homomorphic encryption scheme. For example, in Paillier encryption:

$$E(m_1) \oplus E(m_2) = E(m_1 + m_2 \mod N) \quad (8)$$

– Multiplication (\otimes)

- * **Definition:** The operation that multiplies two encrypted values.
- * **Configuration:** Depends on the specific homomorphic encryption scheme. For example, in ElGamal encryption:

$$E(m_1) \otimes E(m_2) = E(m_1 \cdot m_2 \mod N) \quad (9)$$

• Noise Management

– Noise Growth:

- * **Definition:** During homomorphic operations, the noise in the ciphertext increases.
- * **Configuration:** Noise must be carefully managed to prevent decryption failure.

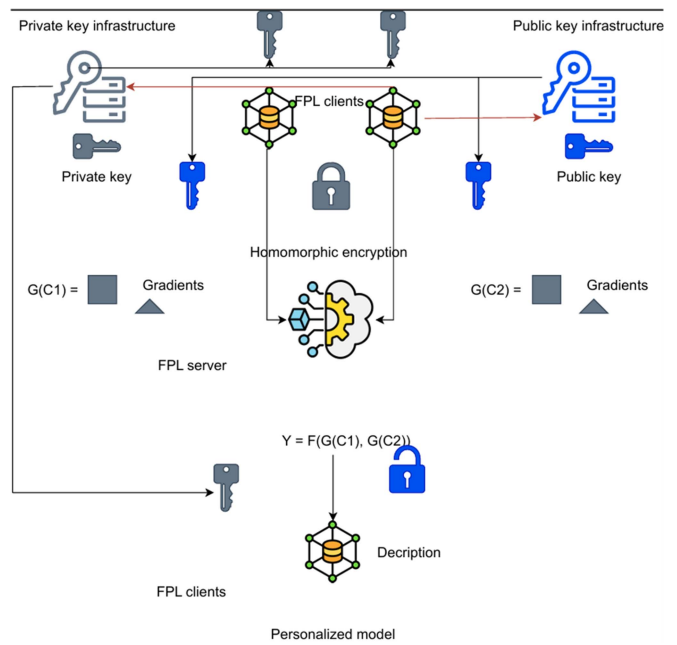


FIGURE 4. Federated personalized learning with Homomorphic Encryption algorithm.

- * **Interpretation:** The level of noise determines how many operations can be performed before the ciphertext becomes undecipherable.

– Bootstrapping

- * **Definition:** A technique to reduce noise in ciphertexts, allowing further computations.
- * **Configuration:** Involves re-encrypting the ciphertext to reset the noise level.
- * **Interpretation:** Essential for fully homomorphic encryption schemes to support unlimited operations.

Fig. 4 illustrates the working procedure of Federated Homomorphic Encryption in detail.

3) FEDERATED PERSONALIZED LEARNING WITH SECURE AGGREGATION AND MULTI-PARTY COMPUTATION

Secure aggregation and secure multi-party computation are cryptographic techniques aimed at enabling collaborative data analysis while preserving the privacy of individual contributions.

Secure aggregation typically involves aggregating sensitive data from multiple parties without revealing individual values [38]. Key parameters include cryptographic protocols such as Additive Homomorphic Encryption (AHE) or secret sharing schemes. AHE allows parties to encrypt their data and send encrypted sums to a trusted aggregator, who can decrypt the aggregate result without knowing the individual contributions. Secret sharing divides data into shares distributed among parties, enabling reconstruction of the aggregate result only when a sufficient number of shares are combined.

Secure multi-party computation extends privacy protection to collaborative computations beyond aggregation [10]. It allows multiple parties to jointly compute a function over their private inputs without revealing those inputs. Mathematical equations in SMPC typically involve protocols like secure function evaluation (SFE) or Yao's garbled circuits. SFE protocols ensure that each party learns only the result of the computation without any knowledge of the other parties' inputs. Yao's garbled circuits encrypt the circuit of the computation, allowing parties to evaluate the circuit while preserving the confidentiality of their inputs. Parameters include security levels based on cryptographic primitives like oblivious transfer and secure comparison protocols, ensuring that computations remain confidential and integrity is maintained across distributed environments.

V. PERFORMANCE ANALYSIS

A. EVALUATION METRICS

We employ a range of performance metrics to evaluate the efficiency of the PPMLFPL models. The model's effectiveness is assessed through a comprehensive set of evaluation metrics, which includes below.

- **Accuracy (A):** Accuracy is a metric that measures the overall correctness of the model's predictions. Mathematically, to compute the accuracy:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

where:

- TP** (True Positives) is the number of correctly predicted positive instances.
- TN** (True Negatives) is the number of correctly predicted negative instances.
- FP** (False Positives) is the number of instances that are actually negative but predicted as positive.
- FN** (False Negatives) is the number of instances that are actually positive but predicted as negative.
- **Precision (P):** Precision represents the proportion of correctly predicted positive instances out of all instances predicted as positive. It indicates the accuracy of positive predictions. Mathematically, to calculate the precision:

$$P = \frac{TP}{TP + FP} \quad (11)$$

- **Recall (R):** Recall measures the ability of a model to capture all positive instances within a dataset. It shows the model's ability to identify positive instances. Mathematically, to compute the recall:

$$R = \frac{TP}{TP + FN} \quad (12)$$

- **F1-score (F):** The F1-score is a commonly used metric in the field of machine learning and classification tasks. It is calculated as the harmonic mean of precision and

TABLE 1. Performance of Evaluation Metrics of Federated Personalized Learning Algorithms. The Evaluation Metrics Include Accuracy, Precision, Recall and f1-Score

Algorithm	Accuracy	Precision	Recall	F1-Score
APFL	93.72	93.75	93.72	93.73
APPLE	97.41	97.41	97.41	97.41
Ditto	82.87	82.9	82.87	82.88
FedALA	67.09	67.12	67.09	67.1
FedFomo	74.31	74.71	74.31	74.51
FedBABU	56.94	57.45	56.94	57.19
FedBN	59.63	59.79	59.63	59.71
FedGC	90.21	90.24	90.21	90.22
FedRep	77.48	77.51	77.48	77.52
FedPAC	84.71	85.11	84.71	84.91
FedPCL	57.75	57.78	57.75	57.76
FedProto	90.09	90.6	90.09	90.34
FML	77.98	78.14	77.98	78.06
HeurFedAMP	75.37	75.88	75.37	75.62
Per-	48.98	49.01	48.98	48.99
FedAvg (FO)				
Per-	52.94	53.34	52.94	53.14
FedAvg (HF)				
pFedMe-PM	91.16	91.32	91.16	91.24

recall. Mathematically, to compute the F1-score:

$$F = \frac{2 \cdot P \cdot R}{P + R} \quad (13)$$

B. RESULTS OF FEDERATED PERSONALIZED LEARNING

This section aims evaluate different federated personalized algorithms that can be used for model personalization.

Fig. 5 represents the accuracy values of various FPL algorithms. The algorithms include APFL, APPLE, Ditto, FedALA, FedFomo, FedBABU, FedBN, FedGC, FedRep, FedPAC, FedPCL, FedProto, FML, HeurFedAMP, Per-FedAvg (FO), Per-FedAvg (HF), and pFedMe-PM. It also illustrates the loss values of various FPL algorithms. The loss values are numerical scores reflecting the error or deviation between the predicted and actual values during the training process.

Table 1 represents the performance of evaluation metrics of different Federated Personalized Learning (FPL) algorithms. The table provides valuable insights into the strengths and weaknesses of different FPL algorithms.

Starting with accuracy, the algorithms show a wide range of values, ranging from as low as 48.98% for Per-FedAvg (FO) to as high as 97.41% for APPLE. This highlights the significant diversity in the ability of the FL algorithms to correctly predict the target classes. Algorithms such as APPLE, APFL, FedGC and pFedMe-PM perform exceptionally well in accurately classifying instances, achieving accuracy scores



FIGURE 5. Accuracy and loss values of federated personalized learning algorithms with changing communication rounds. The blue and green columns indicate the training accuracy and losses, respectively. Whereas, the orange and sky scatter lines indicate the testing accuracy and losses, respectively.

above 90%. On the other hand, FedBABU, Per-FedAvg (FO) and Per-FedAvg (HF) exhibit the lowest accuracy scores, suggesting that they may struggle with capturing the underlying patterns in the data effectively.

Moving on to precision, which indicates the algorithms' ability to minimize false positive predictions, we observe a similar range of values. Again, APPLE stands out as the top-performing algorithm with a precision score of 97.41%, mirroring its high accuracy score. Other algorithms like APFL, FedGC, and pFedMe-PM also demonstrate strong precision performance, indicating their capability to reduce false positives effectively. Conversely, FedBABU, Per-FedAvg (FO) and Per-FedAvg (HF) exhibit the lowest precision scores, implying that they might suffer from a high false positive rate.

Next, analyzing recall, which measures the algorithms' ability to capture all positive instances correctly. Similar to accuracy and precision, APPLE emerges as the leader in recall with a score of 97.41%, showing its proficiency in correctly identifying positive instances. Other high-recall algorithms include APFL, FedGC, and pFedMe-PM, suggesting their effectiveness in capturing positive instances. However, FedBABU, Per-FedAvg (FO) and Per-FedAvg (HF) once again demonstrate the lowest recall scores, indicating their potential limitations in identifying positive instances accurately.

Lastly, examining the F1-score, which combines precision and recall, providing an overall balanced measure of performance. The F1-scores across the algorithms also exhibit a wide range, with APPLE maintaining the highest F1-score of 97.41%. This aligns with its exceptional precision and recall values. Similarly, other top-performing algorithms like APFL, FedGC, and pFedMe-PM also show strong F1-scores, indicating their ability to achieve a balance between precision and recall. Conversely, FedBABU, Per-FedAvg (FO) and Per-FedAvg (HF) once again display the lowest F1-scores, suggesting that they may face challenges in striking a balance between precision and recall.

The Table 1 demonstrates that APPLE performs best across all evaluation metrics, showing high accuracy, precision, recall, and F1-score. It is evident that the performance of the other algorithms varies, with Per-FedAvg (FO) consistently exhibiting the lowest performance.

C. RESULTS OF PRIVACY PRESERVING MACHINE LEARNING WITH FEDERATED PERSONALIZED LEARNING

This section aims to propose and evaluate combined different PPML techniques with previously evaluated best performed federated personalized algorithm APPLE that can be used for model personalization. These algorithms include variations of differential privacy, secure multi-party computation, etc. privacy-preserving techniques.

Table 2 presents the performance of evaluation metrics for different Privacy Preserving Machine Learning with Federated Personalized Learning (PPMLFPL) algorithms. The table provides valuable insights into the strengths and weaknesses of different PPMLFPL algorithms, which is shown in Fig. 6.

TABLE 2. Performance of Evaluation Metrics of Privacy Preserving Federated Personalized Learning Algorithms. The Evaluation Metrics Include Accuracy, Precision, Recall and f1-Score

Algorithm	Accuracy	Precision	Recall	F1-Score
APPLE+DP	97.48	97.51	97.48	97.49
APPLE+HE	99.34	99.34	99.34	99.34
APPLE+SA	97.44	97.47	97.44	97.45
APPLE+SMPC	85.38	85.41	85.38	85.39

Starting with accuracy, the algorithms demonstrate varying levels of performance. The top-performing algorithm is APPLE+HE with an accuracy of 99.34%, which is significantly higher than the other algorithms in the table. This indicates that APPLE+HE excels in correctly predicting the target classes. On the other hand, APPLE+SMPC has the lowest accuracy score of 85.38%, suggesting that it struggles in accurately classifying instances.

Moving on to precision, which measures the algorithms' ability to minimize false positive predictions, APPLE+HE again emerges as the top-performing algorithm with a precision score of 99.34%. This indicates that APPLE+HE has a very low false positive rate, making it highly effective in minimizing false positives. The other algorithms also have relatively high precision scores, ranging from 97.51% for APPLE+DP to 85.41% for APPLE+SMPC.

Next, considering recall, which measures the algorithms' ability to capture all positive instances correctly, we observe that APPLE+HE once again leads with a score of 99.34%. This implies that APPLE+HE is excellent at correctly identifying positive instances. The recall scores for the other algorithms range from 97.48% for APPLE+DP to 85.38% for APPLE+SMPC.

Lastly, examining the F1-score, which combines precision and recall, providing an overall balanced measure of performance, APPLE+HE stands out with an F1-score of 99.34%. This suggests that APPLE+HE achieves a good balance between precision and recall. The other algorithms' F1-scores range from 97.49% for APPLE+DP to 85.39% for APPLE+SMPC.

The Table 2 demonstrates that APPLE+HE performs exceptionally well across all evaluation metrics, showing high accuracy, precision, recall, and F1-score. It is evident that the performance of the other algorithms varies, with APPLE+SMPC consistently exhibiting the lowest performance.

D. POSSIBLE REASON OF WHY APPLE+HE PERFORMS BETTER

APPLE+HE performs better primarily due to the combination of the APPLE algorithm's personalized federated learning capabilities and the robust privacy guarantees provided by HE. HE allows computations to be performed directly on encrypted data, preserving privacy without the need to decrypt the data during the learning process. This reduces the risk

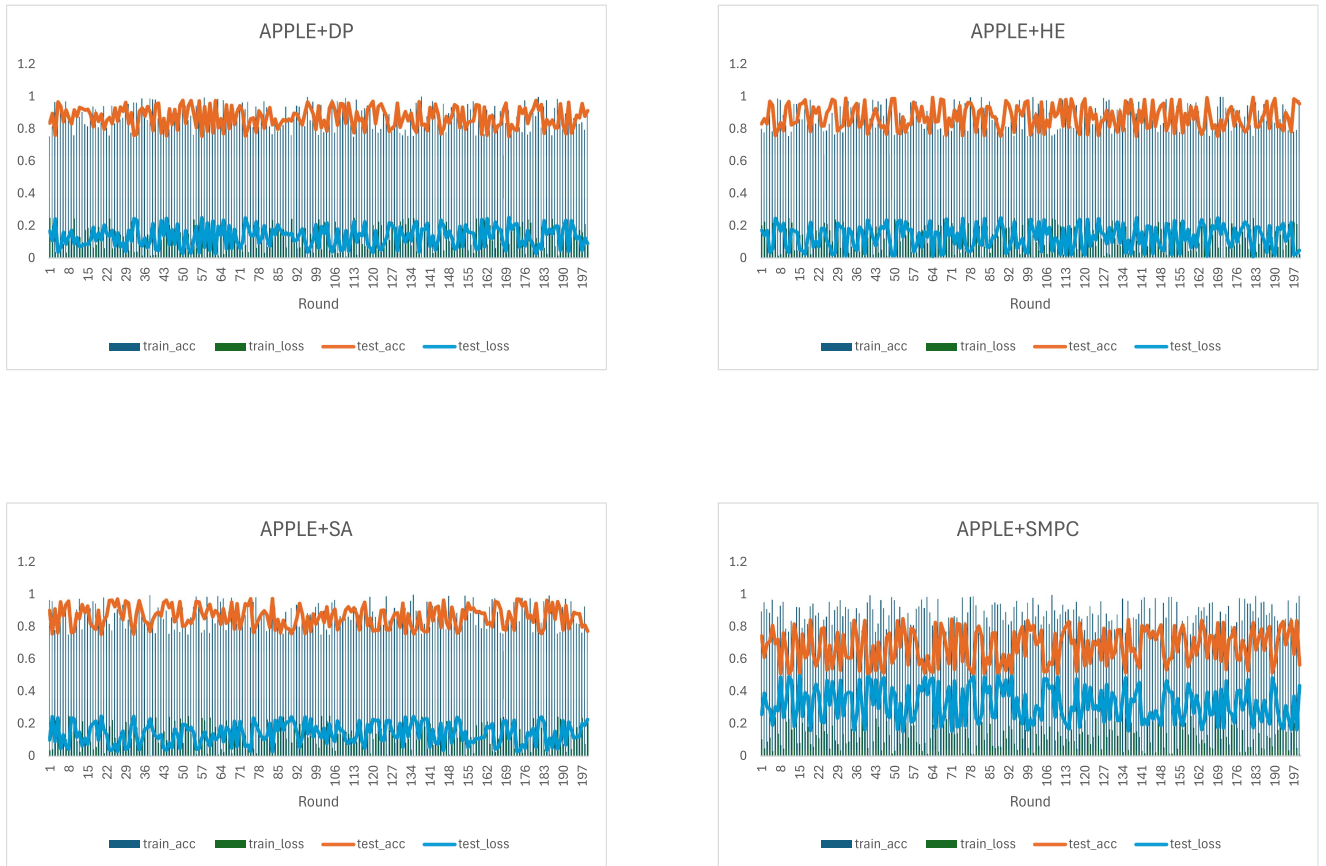


FIGURE 6. Accuracy and loss values of privacy preserving federated personalized learning algorithms with changing communication rounds.

of data breaches and leakage of sensitive information. Moreover, HE ensures that the model updates can be aggregated securely, allowing the APPLE algorithm to effectively utilize the diverse data from different sources without compromising individual privacy. The high accuracy, precision, recall, and F1-score of APPLE+HE suggest that it effectively captures the underlying patterns in the data while maintaining stringent privacy standards, leading to superior performance compared to other privacy-preserving techniques.

E. POTENTIAL TRADE-OFFS AND LIMITATIONS OF APPLE+HE

Despite its superior performance, APPLE+HE comes with certain trade-offs and limitations. The primary trade-off is computational overhead. HE is computationally intensive, significantly increasing the time and resources required for training and inference compared to unencrypted methods. This can lead to longer processing times and higher energy consumption, which may not be feasible for all applications, especially those requiring real-time processing or operating with limited computational resources. Additionally, the complexity of implementing HE can pose challenges in practical deployment, particularly when dealing with large-scale models or datasets. There are also scenarios where the emphasis on privacy preservation can lead to suboptimal learning outcomes. For instance, the noise added by HE to ensure privacy

might degrade the model's accuracy, particularly in applications where high precision is critical. This trade-off between privacy and model performance must be carefully considered, as the benefits of HE in protecting sensitive information may come at the cost of reduced model effectiveness. Moreover, while HE provides strong privacy guarantees, it does not inherently protect against all potential privacy threats, such as inference attacks on model outputs.

VI. FUTURE RESEARCH WORKS AND CONCLUSION

The use of the APPLE+HE algorithm has emerged as a strong recommendation for privacy-preserving machine learning tasks in the artificially generated environment in federated personalized learning settings. HE ensures that computations on encrypted data can be performed without decrypting it, cobining both secure aggregation and secure enclave techniques, maintaining stringent privacy guarantees. Concurrently, adaptive personalization in APPLE optimizes the federated learning process by dynamically adjusting to the heterogeneity of data distributions and computational capacities among silos, thereby enhancing convergence rates and reducing communication overhead. This synergy between advanced cryptographic safeguards and tailored learning protocols ensures robust privacy preservation while maintaining high efficiency in model training and deployment. However,

further research is needed to explore the scalability of APPLE+HE in larger and more diverse virtual environments, as well as its resilience against sophisticated privacy attacks. Future works should also investigate the potential for real-time adaptation to dynamic data distributions within the artificially generated environment. Additionally, evaluating the user experience and acceptance of such privacy-preserving methods will be crucial in ensuring the widespread adoption and success of APPLE+HE in practical applications.

REFERENCES

- [1] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.
- [2] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 22nd ACM Conf. Comput. Commun. Secur.*, 2015, pp. 1310–1321.
- [3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A.Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, 2017, pp. 1273–1282.
- [4] M. Abadi et al., "Deep learning with differential privacy," in *Proc. 2016 ACM Conf. Comput. Commun. Secur.*, 2016, pp. 308–318.
- [5] K. Bonawitz et al., "Practical secure aggregation for federated learning on user-held data," 2016, *arXiv:1611.04482*.
- [6] C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie, "Communication-efficient federated learning via knowledge distillation," *Nature Commun.*, vol. 13, no. 1, 2022, Art. no. 2032.
- [7] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 3557–3568, 2020.
- [8] J. Xu, X. Tong, and S.-L. Huang, "Personalized federated learning with feature alignment and classifier collaboration," 2023, *arXiv:2306.11867*.
- [9] C. Dwork, "Differential Privacy," in *Proc. Int. Colloq. Automata, Lang. Program.*, 2006, pp. 1–12.
- [10] O. Goldreich, "Secure multi-party computation," *Manuscript Preliminary Version*, vol. 78, 1998, Art. no. 110.
- [11] X. Yi, R. Paulet, E. Bertino, X. Yi, R. Paulet, and E. Bertino, *Homomorphic Encryption*. Berlin, Germany: Springer, 2014.
- [12] S. Zhao, Q. Zhang, Y. Qin, W. Feng, and D. Feng, "Sectee: A software-based approach to secure enclave architecture using tee," in *Proc. 2019 ACM Conf. Comput. Commun. Secur.*, 2019, pp. 1723–1740.
- [13] B.-K. Zheng et al., "Scalable and privacy-preserving data sharing based on blockchain," *J. Comput. Sci. Technol.*, vol. 33, pp. 557–567, 2018.
- [14] T. Li, N. Li, J. Zhang, and I. Molloy, "Slicing: A new approach for privacy preserving data publishing," *IEEE Trans. Knowl. data Eng.*, vol. 24, no. 3, pp. 561–574, Mar. 2012.
- [15] V. Costan and S. Devadas, "Intel SGX explained," *Cryptology ePrint Arch.*, 2016.
- [16] R. Buhren, C. Werling, and J.-P. Seifert, "Insecure until proven updated: Analyzing AMD SEV's remote attestation," in *Proc. 2019 ACM Conf. Comput. Commun. Secur.*, 2019, pp. 1087–1099.
- [17] W. Li, Y. Xia, and H. Chen, "Research on arm trustzone," *GetMobile: Mobile Comput. Commun.*, vol. 22, no. 3, pp. 17–22, 2019.
- [18] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, Art. no. 30.
- [19] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," 2021, *arXiv:2102.07623*.
- [20] C. T. Dinh, N. Tran, and J. Nguyen, "Personalized federated learning with Moreau envelopes," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 21394–21405, 2020.
- [21] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and robust federated learning through personalization," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 6357–6368.
- [22] Y. Deng, M. M. Kamani, and M. Mahdavi, "Adaptive personalized federated learning," 2020, *arXiv:2003.13461*.
- [23] M. Zhang, K. Sapra, S. Fidler, S. Yeung, and J. M. Alvarez, "Personalized federated learning with first order model optimization," 2020, *arXiv:2012.08565*.
- [24] Y. Huang et al., "Personalized cross-silo federated learning on non-iid data," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 7865–7873.
- [25] X.-C. Li, D.-C. Zhan, Y. Shao, B. Li, and S. Song, "Fedphp: Federated personalization with inherited private models," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2021, pp. 587–602.
- [26] J. Luo and S. Wu, "Adapt to adaptation: Learning personalization for cross-silo federated learning," in *Proc. IJCAI Conf.*, 2022, Art. no. 2166.
- [27] J. Zhang et al., "Fedala: Adaptive local aggregation for personalized federated learning," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 11237–11244.
- [28] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, "Federated learning with personalization layers," 2019, *arXiv:1912.00818*.
- [29] P. P. Liang et al., "Think locally, act globally: Federated learning with local and global representations," 2020, *arXiv:2001.01523*.
- [30] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "Exploiting shared representations for personalized federated learning," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 2089–2099.
- [31] H.-Y. Chen and W.-L. Chao, "On bridging generic and personalized federated learning for image classification," 2021, *arXiv:2107.00778*.
- [32] J. Oh, S. Kim, and S.-Y. Yun, "Fedbabu: Towards enhanced representation for federated image classification," 2021, *arXiv:2106.06042*.
- [33] Y. Niu and W. Deng, "Federated learning for face recognition with gradient correction," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 2007–2022.
- [34] H. Seo, J. Park, S. Oh, M. Bennis, and S.-L. Kim, "16 Federated knowledge distillation," *Machine Learning and Wireless Communications*. Cambridge Univ. Press, pp. 457–485.
- [35] T. Shen et al., "Federated mutual learning," 2020, *arXiv:2006.16765*.
- [36] Y. Tan et al., "Fedproto: Federated prototype learning across heterogeneous clients," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 8432–8440.
- [37] Y. Tan, G. Long, J. Ma, L. Liu, T. Zhou, and J. Jiang, "Federated learning from pre-trained models: A contrastive learning approach," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 19332–19344, 2022.
- [38] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in *Proc. 2017 ACM Conf. Comput. Commun. Secur.*, 2017, pp. 1175–1191.