Fairness in Federated Learning via Collaborative Aggregation: an Empirical Study

Wayne Lemieux , Mitra Hassani

Abstract

Current training methods for deep learning models often struggle with overfitting when data is kept on individual devices (Standalone approach). This leads to models that perform poorly on unseen data. Distributed Learning (DL) and Federated Learning (FL) tackle this issue by combining updates from multiple devices through a central server. However, these approaches often ignore fairness in how models are shared. In other words, everyone gets the same model regardless of how much they contribute. Our research focuses on fairness in FL. We propose a new system called Collaborative Fair Federated Learning (CFFL) that uses a reputation system to encourage participants to develop different models. This method promotes fairness without hurting the model's ability to make accurate predictions. Tests on standard datasets show that CFFL is both fairer and performs just as well as DL, while significantly outperforming Standalone methods.

1 Introduction

Training powerful deep learning models requires massive datasets and processing power, making it a challenge for individual participants. Even with a local dataset, building a complex model can lead to poor performance on unseen data. Federated Learning (FL) tackles this by enabling collaborative training on private datasets across multiple devices. By combining diverse data from many participants, FL aims to create a more generalizable global model [1].

However, current FL approaches have a fairness issue. All participants receive the same model updates, regardless of their contribution. This is problematic because participants contribute unevenly due to differences in data quality and quantity. Some participants' data significantly improves the model, while others might even hinder it. Imagine several banks collaborating on a credit score predictor for small businesses. Larger banks with more high-quality data might be hesitant to share it freely, fearing it could give smaller banks a competitive edge [1–4].

**Furthermore, existing FL can't distinguish between high and low contributors, creating a vulnerability to "free-riders" who benefit without contributing much. This lack of fairness hinders FL adoption, especially in collaborations between different organizations. It not only discourages a healthy FL ecosystem but also limits widespread implementation. While fairness research often focuses on mitigating bias in model outputs, ensuring fair treatment based on contribution remains an open challenge [5].

Here, we address this challenge with Collaborative Fair Federated Learning (CFFL). Unlike existing approaches that rely on external incentives, CFFL modifies the FL process itself. This ensures participants receive models with performance proportional to their contributions, replacing the one-size-fits-all model. CFFL leverages a reputation mechanism that tracks and updates participant contributions throughout the learning process.

This approach is particularly valuable for business-to-business (H2B) federated learning, especially in sensitive sectors like finance or healthcare, where fairness is critical [6]. Our goal is to establish collaborative fairness in FL by dynamically adjusting model performance based on contribution [7, 8].. Experiments show that CFFL achieves the highest fairness level, with the most contributive participant's accuracy matching the Distributed learning approach and exceeding the Standalone method. We'll use "Distributed" and "Federated" interchangeably throughout the paper.

2 Related Work

This section reviews relevant literature on fairness in FL to situate our research within the existing landscape.

2.1 Incentive Schemes for Collaborative Fairness

One prominent approach to fostering collaboration and fairness among FL participants involves incentive schemes [3]. These schemes reward participants based on their contributions. Egalitarian profit-sharing, like equal division [9], distributes the total reward equally across participants in a round. Individual profit-sharing [9] assigns rewards based on each participant's contribution. The Labour Union game [10] and Fair-value game [10] schemes consider a participant's impact on the collective utility and their joining sequence, respectively. However, the Shapley game scheme [10], while aiming for fairness, becomes computationally expensive for large-scale FL due to its exponential complexity.

2.2 Rewarding Model Improvement in Gradient-based FL

For gradient-based FL, where gradients are valuable data, output agreement-based rewards are challenging. Therefore, model accuracy becomes the primary criterion for designing rewards. Richardson et al. [11] propose a scheme that compensates for marginal improvements in model updates while addressing potential overestimation through a correction model. This approach encourages early participation and predictable budget allocation for achieving target accuracy levels.

2.3 Multi-Objective Optimization for Fairness

Recent works like [12–14] explore FL as a multi-objective optimization problem, aiming to balance multiple objectives, including fairness, during FL training. Yu et al. [6] introduce a joint objective optimization approach that considers participant contributions, costs, and waiting time to achieve broader fairness in reward distribution within FL.

3 The CFFL Framework

3.1 Collaborative Fairness

Our proposed framework presents a novel approach to Federated Learning (FL), departing from existing methods by tailoring the FL model to each participant's contributions, thus introducing a new paradigm. This alternative framework aims to enhance collaborative fairness, a concept we define and elaborate on below.

Definition 1. *Collaborative fairness in a federated system refers to the principle that participants with higher contributions should receive local models that outperform those provided to participants with lower contributions. Mathematically, fairness is quantified by the correlation coefficient between participants' contributions and their respective final model accuracies.*

3.2 Fairness

In our Collaborative Fair Federated Learning (CFFL) approach, we modify the FL process to enable participants to download only aggregated updates based on their reputations. The server maintains a reputation list for all participants, updating it according to the quality of gradients uploaded in each communication round. Participants upload gradients according to a selected upload rate, denoted as θ_u. For instance, if $\theta_u = 1$, all gradients are uploaded; if $\theta_u = 0.1$, only 10% of gradients are uploaded. The server evaluates participants' validation accuracy based on the uploaded gradients.

If $\theta_u = 1$, the server computes the validation accuracy $vacc_j$ using participant j's entire model $\boldsymbol{w}_j$. If $\theta_u \neq 1$, participant j's uploaded gradients $\Delta(\boldsymbol{w}_j)^S$ are integrated into an auxiliary model $\boldsymbol{w}_g$ to compute $vacc_j$. Importantly, $\boldsymbol{w}_g$ is maintained by the server for gradient aggregation and reputation calculation, and its parameters are not broadcasted to individual participants as in standard FL systems.

The server then normalizes $vacc_j$ and calculates the reputation c_j of participant j in each communication round using a $sinh(\alpha)$ function (Eq. (**??**)), where x is the normalized $vacc_j$. The higher x, the more informative participant j's uploaded gradients are. The $sinh(\alpha)$ function, with α as the punishment factor, distinguishes reputations based on gradient informativeness. The server iteratively updates each participant's reputation based on the calculated and historical reputations.

Higher reputation leads to more allocated aggregated gradients, serving as a reward in each communication round. The detailed implementation of CFFL involves participants uploading clipped gradients to the server, with reputations updated based on gradient performance on a validation set, determining the number of aggregated updates allocated to each participant. Gradient clipping is used to mitigate noise from abnormal examples or outliers.

3.3 Measuring Fairness

In this study, collaborative fairness is quantified through the correlation coefficient between participant contributions and rewards. The X-axis represents the test accuracies of standalone models, characterizing participants' individual learning capabilities on their local datasets. The Y-axis represents the test accuracies of final models received by the participants.

Participants with higher standalone accuracies are considered to contribute more. Thus, the X-axis is expressed as $\boldsymbol{x} = \{sacc_1, \cdots, sacc_n\}$, where $sacc_j$ denotes the standalone model accuracy of participant j. Similarly, the Y-axis is expressed as $\boldsymbol{y} = \{acc_1, \cdots, acc_n\}$, where acc_j represents the final model accuracy of participant j after collaboration.

As the Y-axis measures the respective model performance of different participants after collaboration, a positive correlation with the X-axis is expected for a fair measure. Collaborative fairness is formally quantified using the correlation coefficient in Equation 1:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{1}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of $\boldsymbol{x}$ and $\boldsymbol{y}$, and s_x and s_y are the corrected standard deviations. The fairness coefficient ranges from -1 to 1, with higher values indicating better fairness. A negative coefficient implies poor fairness.

4 Experimental Evaluation

4.1 Datasets

We conducted experiments on two benchmark datasets: the MNIST dataset for handwritten digit recognition and the Adult Census dataset.

MNIST Dataset: This dataset consists of 60,000 training examples and 10,000 test examples, sourced from [MNIST](http://yann.lecun.com/exdb/mnist/).

Adult Census Dataset: This dataset, commonly used to predict whether an individual earns over 50,000 annually (binary), contains a total of 48,843 records. We manually balanced the dataset to include 11,687 records above and below the 50,000 threshold, resulting in a total of 23,374 records. The dataset attributes include age, race, education level, marital status, occupation, among others. Data is sourced from the [Adult Census dataset](http://archive.ics.uci.edu/ml/datasets/Adult).

For both datasets, we conducted an 80-20 train-test split and randomly selected 10

4.2 Baselines

We demonstrate the effectiveness of our proposed CFFL framework through comparison with the following frameworks. *Standalone* framework and two representative *Distributed* baselines.: FedAvg [1] and DSSGD [15]. for FL.

We showcase the effectiveness of our proposed Collaborative Fair Federated Learning (CFFL) framework by comparing it with the following frameworks:

Standalone Framework: This baseline represents the traditional approach where participants train their models independently without collaboration.

Distributed Baselines: FedAvg (Federated Averaging): This is a widely used federated learning algorithm introduced by McMahan et al. (2017) [1]. FedAvg aggregates the model updates from all participants to create a global model.

DSSGD (Distributed Stochastic Gradient Descent): This framework, proposed by Shokri et al. (2015) [15], utilizes a round-robin communication protocol where participants upload parameter updates and download parameters in sequence, allowing for model aggregation.

These baselines provide a benchmark against which we evaluate the performance and effectiveness of our CFFL framework in terms of fairness, utility, and convergence behavior in federated learning settings.

Furthermore, we investigate different upload rates $\theta_u = 0.1$ and $\theta_u = 1$ [15], where gradients are uploaded according to the "largest values" criterion when $\theta_u = 0.1$. The rationale behind introducing an upload rate less than 1 is to reduce overfitting and to save communication overhead. We observe a similar practice by the FedAvg algorithm that in a large-scale FL system, it may randomly select a subset of participants in each communication round for parameter update and sharing, in order to reduce the risk of overfitting and to reduce communication cost.

4.3 Experimental Setup

To investigate data distribution heterogeneity, we conducted experiments using the MNIST dataset where we varied the number of classes in the data of each participant. We distributed classes in a linspace manner, such that the first participant owned data from only one class, while the last participant owned data from all ten classes. Specifically, for a total of 10 classes and 5 participants, we simulated the first participant having data from only one class, while the last participant had data from all ten classes. Each participant retained the same number of examples, i.e., 600 examples.

Model and Hyper-Parameters: For the MNIST "Imbalanced Data Size" experiment, we utilized a two-layer fully connected neural network with 128 and 64 units respectively. The hyperparameters included a local epoch of $E = 2$, local batch size of $B = 16$, and local learning rate of $lr = 0.15$ for $P = 5$ participants and $lr = 0.25$ for $P = \{10, 20\}$ participants. Exponential decay with $\gamma = 0.977$ was applied, along with gradient clipping between $[-0.01, 0.01]$. A total of 30 communication rounds were executed. For the MNIST "Imbalanced Class Numbers" experiment, the same neural network architecture was employed with local epochs of $E = 1$, local batch size of $B = 16$, and local learning rate of $lr = 0.2$ for $P = \{5, 10, 20\}$ participants. Exponential decay with $\gamma = 0.966$ and gradient clipping between $[-0.01, 0.01]$ were also utilized, with a total of 50 communication rounds. For the Adult dataset, a single-layer fully connected neural network with 32 units was utilized, with hyperparameters including local epochs of $E = 2$, local batch size of $B = 16$, and local learning rate of $lr = 0.03$. Exponential decay with $\gamma = 0.966$ and gradient clipping between $[-0.01, 0.01]$ were also employed, with a total of 30 communication rounds.

To mitigate the impact of different initializations and ensure convergence, the same model parameter $\boldsymbol{w}_0$ was initialized for all participants and the server $\boldsymbol{w}_g$. The reputation threshold was empirically set via grid search as $c_{th} = \frac{1}{|R|} * \frac{1}{3}$ for imbalanced data size and $c_{th} = \frac{1}{|R|} * \frac{1}{6}$ for imbalanced class numbers, where $|R|$ is the number of participants with reputations higher than the threshold. The punishment factor was empirically chosen as $\alpha = 5$. Stochastic Gradient Descent (SGD) was used as the optimization technique throughout.

Communication Protocol: To enable the calculation of fairness using the Pearson coefficient, we adopted the round-robin communication protocol for DSSGD and FedAvg. In each communication round, participants uploaded parameter updates and downloaded parameters in sequence, resulting in models with insignificant performance differences. This facilitated the use of test accuracies for the calculation of fairness. FedAvg was not explicitly modified to accommodate fairness calculations, and was solely considered as a performance comparison baseline.

4.4 Experimental Results

Fairness comparison. Table 1 provides the calculated fairness values for DSSGD, FedAvg, and CFFL across the MNIST and Adult datasets, considering varying participant numbers from $\{5, 10, 20\}$, different pretraining statuses from $\{1, 0\}$, and different upload rates θ_u from $\{0.1, 1\}$. The obtained high fairness values, some approaching the theoretical limit of 1.0, affirm that CFFL achieves commendable fairness. This observation aligns with our notion of fairness, where participants with higher contributions are rewarded with better-performing models.

Moreover, the inclusion of pretraining (pretrain=1) tends to result in slightly higher fairness compared to scenarios without pretraining (pretrain=0). This can be attributed to the individual pretraining of 5 epochs before collaborative learning commences, which helps participants' models to advance closer to their respective optima. It's important to note that pretraining is exclusively conducted for CFFL.

Additionally, we observe that both DSSGD and FedAvg yield significantly lower fairness values compared to our CFFL. This outcome is anticipated, as neither the communication protocol nor the learning algorithm in these frameworks incorporates the concept of fairness. Overall, the results presented in Table 1 underscore the effectiveness of CFFL in achieving high levels of fairness, further validating the utility of our proposed fairness mechanism.

Table 1: Fairness [%] of DSSGD, FedAvg and CFFL under varying participant number settings (P-k), pretraining status and upload rate θ_u. .

Dataset	MNIST							Adult						
Framework	FedAvg	DSSGD		CFFL				FedAvg	DSSGD		CFFL			
Pretrain	NA	NA		1		0		NA	NA		1		0	
θ_u	NA	0.1	1	0.1	1	0.1	1	NA	0.1	1	0.1	1	0.1	1
P5	3.08	90.72	84.61	99.63	98.66	**99.76**	99.02	-3.33	15.61	35.71	98.50	97.75	98.44	**99.37**
P10	-50.47	-78.18	90.67	97.90	97.30	98.55	**98.74**	44.27	62.30	56.60	88.00	**93.07**	92.00	91.95
P20	60.41	-81.77	80.45	**99.23**	96.28	98.52	98.51	-34.32	60.30	58.01	**84.41**	82.46	80.56	79.52

Table 2: Maximum Accuracy [%] over MNIST and Adult of varying participant number settings, achieved by DSSGD, FedAvg, *Standalone* framework, and our CFFL ($\theta_u = 0.1$, where CFFL* denotes CFFL with pretraining).

Framework	MNIST			Adult		
	P5	P10	P20	P5	P10	P20
DSSGD	93.28	94.20	82.36	81.94	82.78	82.07
FedAvg	**93.62**	**95.32**	**96.26**	**82.58**	**83.14**	**83.16**
Standalone	90.30	90.88	90.64	81.93	82.31	82.07
CFFL	91.83	93.00	93.25	81.96	82.63	82.72
*CFFL**	91.85	92.85	93.34	81.89	82.63	82.63

Accuracy comparison. Table 2 presents the corresponding accuracies on the MNIST and Adult datasets for scenarios involving $\{5, 10, 20\}$ participants with an upload rate of $\theta_u = 0.1$. We report the highest accuracy achieved among the participants, as Collaborative Fair Federated Learning (CFFL) allows participants to converge to different final models. Therefore, we expect the most contributive participant to obtain a model with the highest accuracy, which is comparable to both the Distributed frameworks.

For the Standalone framework, we display the accuracy of the same participant. It is evident from the table that CFFL consistently achieves comparable accuracy to DSSGD and FedAvg, and consistently outperforms the Standalone framework. For instance, in the case of MNIST with 20 participants, our CFFL (CFFL*) achieves a test accuracy of 92.11 (92.55)%, which is higher than the Standalone framework's accuracy of 91.41%, and slightly lower than FedAvg's accuracy of 97.12%. The lower accuracy of DSSGD in this setting can be attributed to its higher instability and fluctuations during training.

The fairness results presented in Table 1, coupled with the accuracy results in Table 2, demonstrate that our proposed CFFL achieves reasonable fairness without significantly compromising accuracy across various settings.

Individual model performance. To evaluate the impact of our Collaborative Fair Federated Learning (CFFL) framework on individual convergence, we plotted the test accuracy of each participant for both the Standalone framework and CFFL. We conducted experiments with upload rates of $\{0.1, 1\}$ and compared results with and without pretraining over the MNIST dataset across 30 communication rounds.

Our findings consistently demonstrate that CFFL yields superior accuracy compared to the standalone model of any participant. This observation reaffirms that CFFL effectively guides participants to converge to different local models, each of which outperforms their standalone counterparts. This outcome aligns with our claims of offering both fairness and utility. It's worth noting that CFFL with pretraining may take slightly longer (less than 5 additional rounds) for all participants to converge compared to CFFL without pretraining. This trend is observed across both the MNIST and Adult datasets. Additionally, slight fluctuations are observed at the beginning of training. This phenomenon can be attributed to the fact that participants are allocated different aggregated updates from the server. Overall, the convergence curves for CFFL with and without pretraining follow similar trends, confirming that pretraining does not significantly alter the overall convergence behavior. However, pretraining does provide relatively better fairness in most cases, reinforcing the benefits of incorporating pretraining into the CFFL framework.

These observations hold true for both the MNIST and Adult datasets, as depicted in the convergence plots provided.

In the scenario of imbalanced class numbers, we observe the individual model accuracies in both the Standalone framework and our Collaborative Fair Federated Learning (CFFL). Notably, all participants achieve higher accuracies in CFFL compared to their standalone counterparts. This improvement highlights the effectiveness of CFFL in enhancing model performance. Similar to the imbalanced data size scenario, we notice that all participants in CFFL converge to different final models, leading to more pronounced accuracy gaps among them. This outcome contributes to higher levels of fairness in the system. Furthermore, we observe that the convergence process takes longer when there are more participants involved in the system. This observation underscores the impact of participant numbers on convergence time within the federated learning framework.

5 Discussions

In the realm of Federated Learning (FL), the existence of free-riders—participants aiming to exploit the global model without genuine contributions—poses a significant challenge. These individuals may attempt to mask their lack of contribution by uploading random or noisy updates. Traditional FL systems lack specific mechanisms to counteract this behavior, allowing free-riders to exploit the benefits of the global model without substantial investment. In contrast, our Collaborative Fair Federated Learning (CFFL) framework possesses the capability to automatically detect and isolate free-riders.

This is accomplished through the empirical assessment of utility on the validation set, where random or noisy gradients typically yield low utility scores. As collaborative training progresses, free-riders experience a gradual decline in their reputations, ultimately leading to their isolation from the system once their reputations dip below the predefined threshold.

In our supplementary experiments, including a scenario where a free-rider consistently uploads random values as gradients, CFFL consistently identifies and isolates the free-rider during the initial stages of collaboration. Remarkably, this isolation occurs without compromising accuracy or convergence, underscoring the robustness of our framework against free-rider exploitation.

The selection of a reputation threshold, denoted as c_{th}, plays a crucial role in managing participant reputations and detecting potential free-riders within the federated system. This threshold serves as a lower bound on participant reputations and aids in maintaining fairness and accuracy.

Choosing an appropriate c_{th} value is essential, as it directly impacts the balance between fairness and accuracy. Setting a very small c_{th} might allow low-contribution participants to evade detection and remain integrated into the federated system. Conversely, a too large c_{th} could result in the unjust isolation of numerous participants.

In our experiments, we empirically determine suitable c_{th} values tailored to different scenarios. For example, when dealing with imbalanced data sizes, we set c_{th} to $\frac{1}{|R|} * \frac{1}{3}$. In scenarios with imbalanced class numbers, we adjust c_{th} to $\frac{1}{|R|} * \frac{1}{6}$. These selected values strike a delicate balance between effectively identifying free-riders and preserving the integrity of the federated system.

6 Conclusion

The study presented here marks a pioneering effort in exploring collaborative fairness within Federated Learning (FL). By introducing modifications to FL that steer participants towards distinct models, the proposed Collaborative Fair Federated Learning (CFFL) framework represents a significant advancement. This framework integrates a novel reputation mechanism that influences participant rewards based on their empirical model performance across communication rounds on a validation set.

Experimental findings underscore the efficacy of CFFL, demonstrating comparable accuracy to two Distributed frameworks while consistently outperforming the Standalone framework in terms of both fairness and utility.

Numerous avenues for future research emerge from this work. One promising direction involves quantifying fairness in more complex settings characterized by diverse data distributions. Furthermore, extending the application of the CFFL framework to various domains such as finance, biomedical science, speech processing, and natural language processing could yield valuable insights.

Additionally, there is a compelling need for a systematic exploration of integrating robustness with fairness. Investigating the robustness of CFFL under different participant assumptions, including adversarial scenarios, holds potential for developing a more resilient federated learning framework.

The extensive real-world applications of our system underscore the importance of further exploration and research in this area.

References

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, 2017, pp. 1273–1282.

[2] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al., "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.

[3] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[4] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith, "Federated learning: Challenges, methods, and future directions," *CoRR, arXiv:1908.07873*, 2019.

[5] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu, *Federated Learning*, Morgan & Claypool Publishers, 2019.

[6] Han Yu, Zelei Liu, Yang Liu, Tianjian Chen, Mingshu Cong, Xi Weng, Dusit Niyato, and Qiang Yang, "A fairness-aware incentive scheme for federated learning," in *Proceedings of the 3rd AAAI/ACM Conference on AI, Ethics, and Society (AIES-20)*, 2020, pp. 393–399.

[7] Lingjuan Lyu, Jiangshan Yu, Karthik Nandakumar, Yitong Li, Xingjun Ma, Jiong Jin, Han Yu, and Kee Siong Ng, "Towards fair and privacy-preserving federated deep models," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 11, pp. 2524–2541, 2020.

[8] Lingjuan Lyu, Yitong Li, Karthik Nandakumar, Jiangshan Yu, and Xingjun Ma, "How to democratise and protect ai: Fair and differentially private decentralised deep learning," *IEEE Transactions on Dependable and Secure Computing*, 2020.

[9] Shuo Yang, Fan Wu, Shaojie Tang, Xiaofeng Gao, Bo Yang, and Guihai Chen, "On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 832–847, 2017.

[10] Sreenivas Gollapudi, Kostas Kollias, Debmalya Panigrahi, and Venetia Pliatsika, "Profit sharing and efficiency in utility games," in *ESA*, 2017, pp. 1–16.

[11] Adam Richardson, Aris Filos-Ratsikas, and Boi Faltings, "Rewarding high-quality data via influence functions," *arXiv preprint arXiv:1908.11598*, 2019.

[12] Zeou Hu, Kiarash Shaloudegi, Guojun Zhang, and Yaoliang Yu, "Federated learning meets multi-objective optimization," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2039–2051, 2022.

[13] Shayan Mohajer Hamidi and En-Hui Yang, "Adafed: Fair federated learning via adaptive common descent direction," *arXiv preprint arXiv:2401.04993*, 2024.

[14] Shayan Mohajer Hamidi and Oussama Damen, "Fair wireless federated learning through the identification of a common descent direction," *IEEE Communications Letters*, pp. 1–1, 2024.

[15] Reza Shokri and Vitaly Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2015, pp. 1310–1321.