Challenges and Opportunities in Federated Unlearning

Hyejun Jeong, Shiqing Ma, Amir Houmansadr

University of Massachusetts Amherst

Outline

- Preliminary: RTBF, MU, FL
- Challenges in Federated Unlearning
- Federated Unlearning
  - Who unlearns
  - What dataset
  - Learning config
  - Research implication
- Evaluation Objectives and Metric
- Insights and Future Research Direction

RTBF, MU

- The Right To Be Forgotten (RTBF)
  - An individual can request to eliminate their information and **the influence on a trained model** if they withdraw their consent.

- Machine Unlearning (MU)
  - Naïve approach: retrain the model from scratch, excluding the data to forget (retrain)

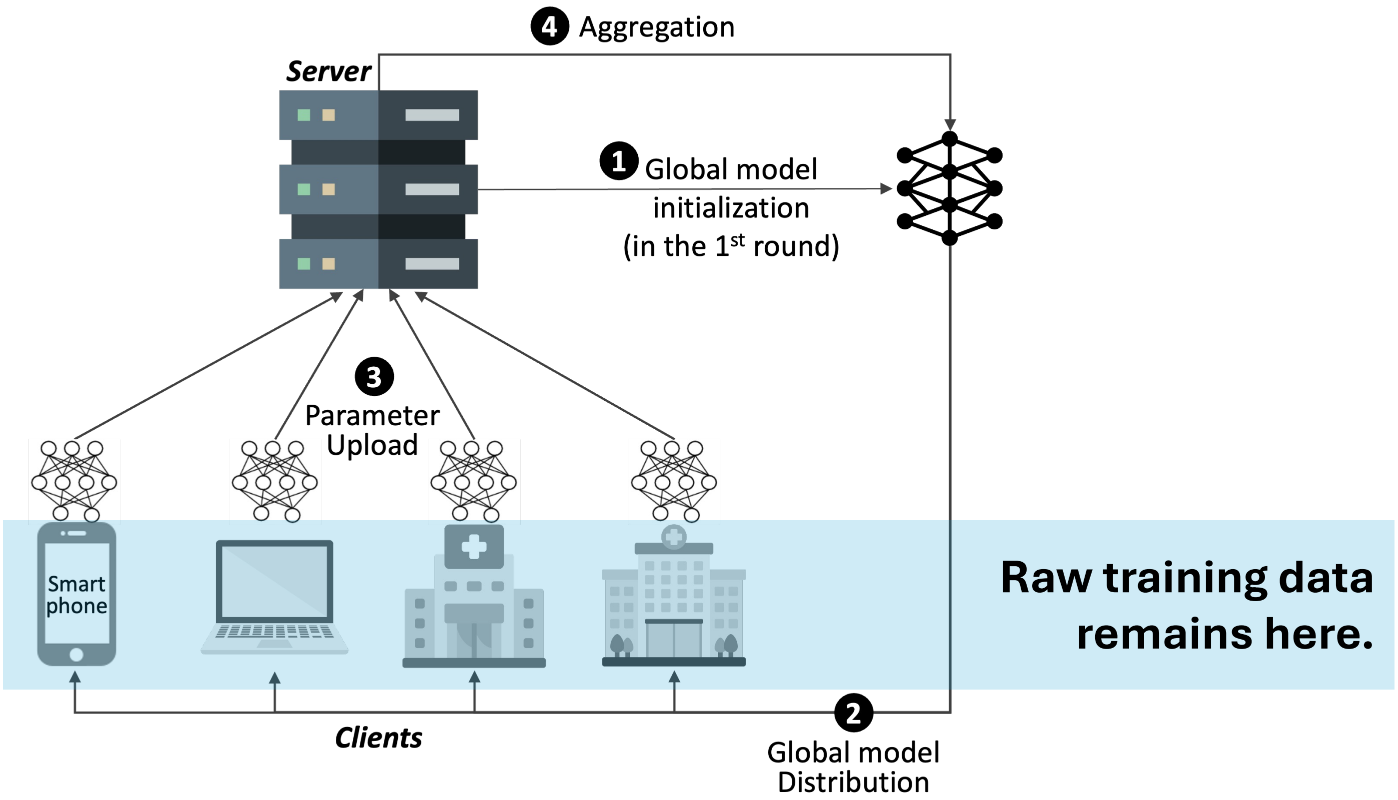
    ➔ **Infeasible** due to overhead

    - time, memory, and resource consumption

  - Efficiently remove the target's influence from the trained model
    - Data-driven: partition, obfuscation, augmentation
    - Model manipulation: shifting, pruning, replacement

Federated Learning

- A distributed machine learning framework preserving data privacy

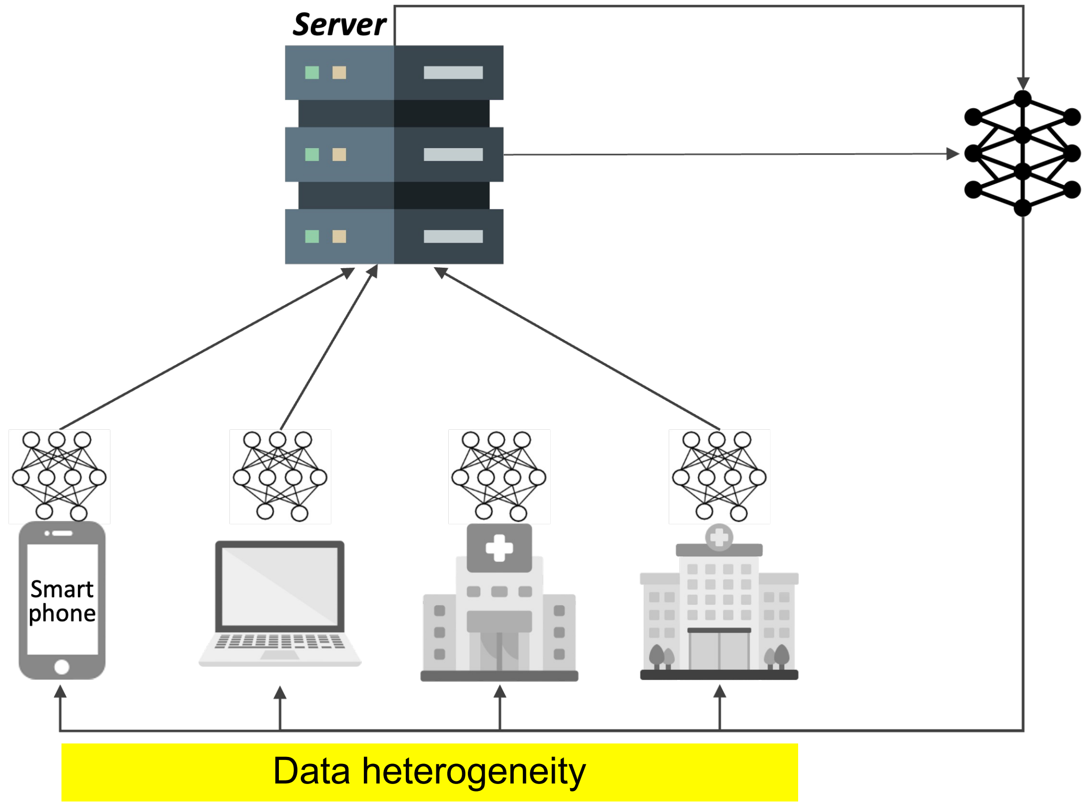


Federated Learning

Challenges in Federated Unlearning

University *of* Massachusetts Amherst

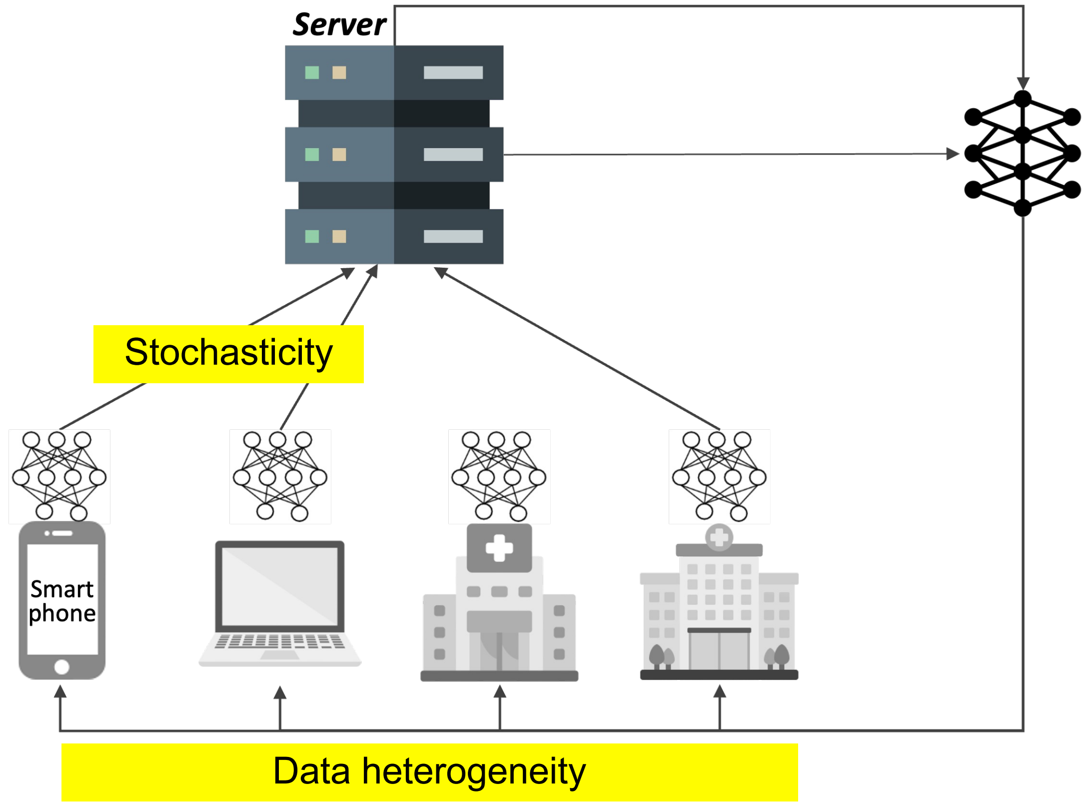
- Data heterogeneity



Challenges in Federated Unlearning

University of Massachusetts Amherst

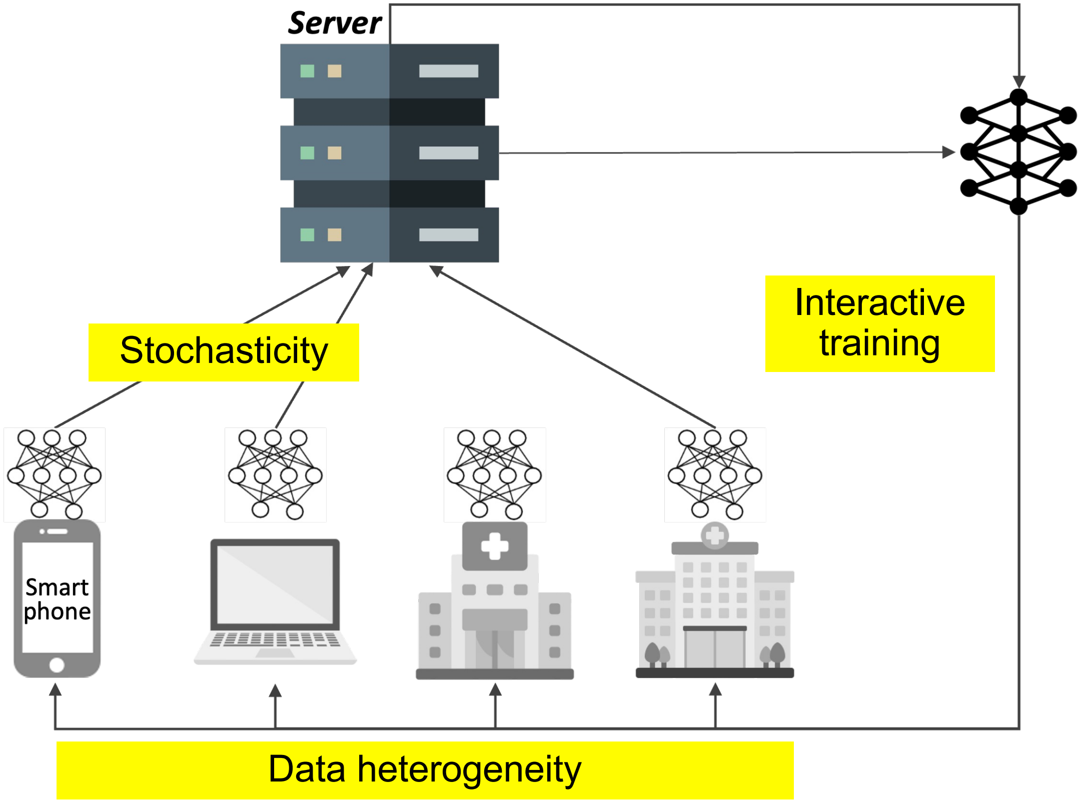
- Data heterogeneity
- Stochasticity of client selection



Challenges in Federated Unlearning

University of Massachusetts Amherst

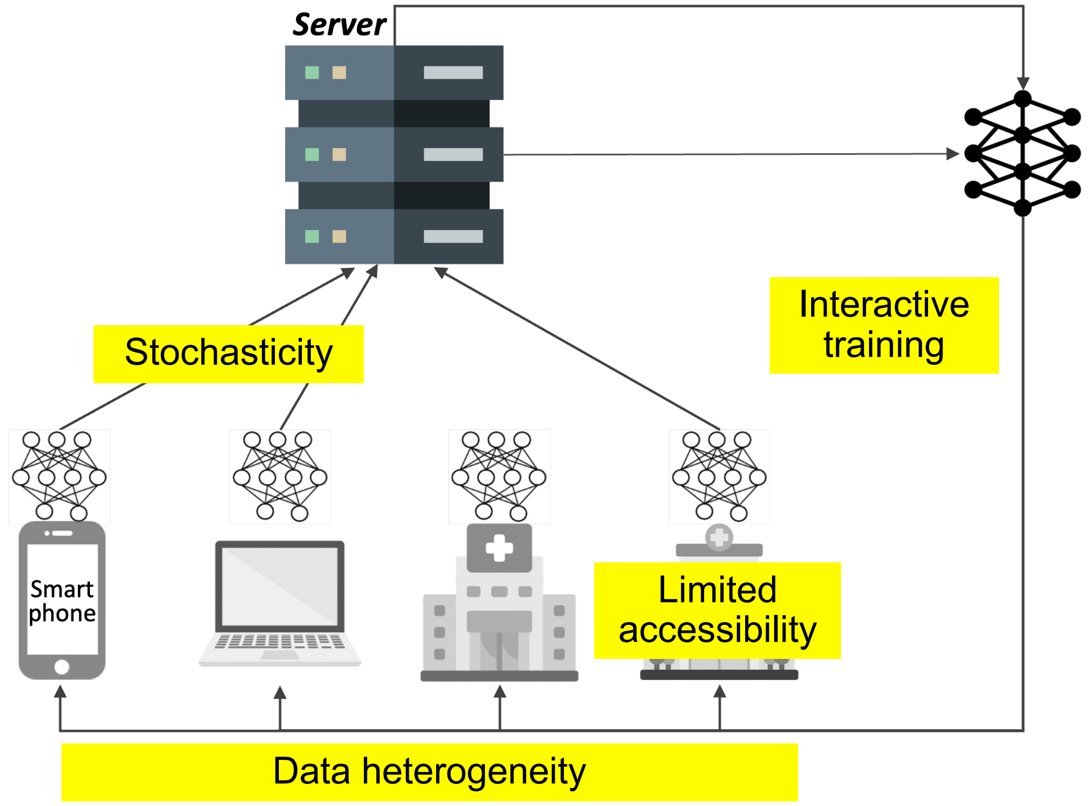
- Data heterogeneity
- Stochasticity of client selection
- Interactive training



Challenges in Federated Unlearning

- Data heterogeneity
- Stochasticity of client selection
- Interactive training
- Limited accessibility

Unlearner	Global	Own local	All local	Raw data
Server	✓		✓	
Target client	✓	✓		✓
Remaining clients	✓	✓		

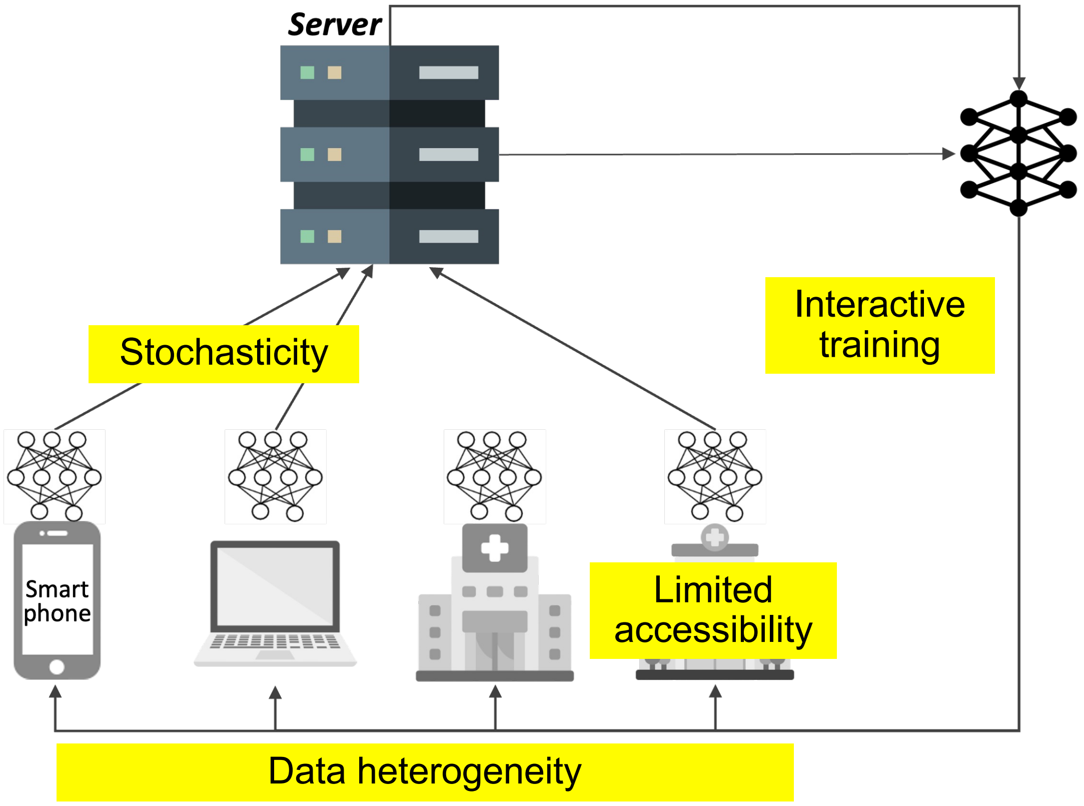


Challenges in Federated Unlearning

University of Massachusetts Amherst

- Data heterogeneity
- Stochasticity of client selection
- Interactive training
- Limited accessibility

Unlearner	Global	Own local	All local	Raw data
Server	✓		✓	
Target client	✓	✓		✓
Remaining clients	✓	✓		



- ***Unlearning techniques in centralized settings are not trivially applicable!***

Federated Unlearning

- We reviewed 44 Federated Unlearning papers.
  - System models
    - Who unlearns?
    - What data distribution?
    - What dataset?
    - Learning config?
    - Research implications?
  - Unlearning techniques
  - Evaluation metrics

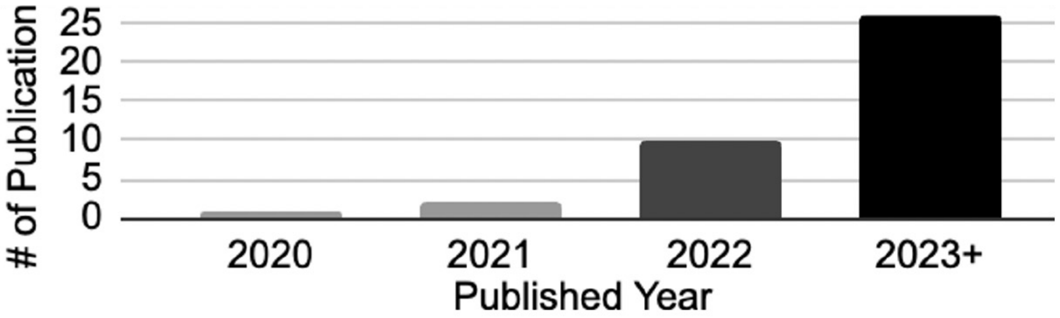


Figure 1: Number of Federated Unlearning Publications.

Federated Unlearning

University of Massachusetts Amherst

Removal Request

Federated Unlearning

FL Model

Target Removal

- Sample
- Class
- Client
- Feature

Influence Removal

- Historical info
- Gradient manipulation
- Loss function estimation
- Knowledge distillation
- Multi-task learning
- Reverse training
- Clustering

Performance Recovery

- Post training
- Fine-tuning
- Gradient manipulation
- Regularization
- Knowledge Distillation

Evaluation Metric

- Performance metric
- Parameter difference
- Indiscrimination quality
- Time taken
- Memory used

Unlearned Model

Who Unlearns under What Data Distribution?

Unlearner	Global	Own local	All local	Raw data
Server	✓		✓	
Target client	✓	✓		✓
Remaining clients	✓	✓		

Who Unlearns?

- Available knowledge varies depending on who unlearns.
- What if a target request removal and leave?

Ref.	Unlearner			Data Dist.	NIID sim.
	Server	Target	Remain		
RevFRF[37]	●			n/d	n/d
Exact-Fun[68]		●	●	Non-IID	random
FATS[56]	●			Non-IID	Dirichlet
Shao et al.[52]	●			Non-IID	unique
Wang et al.[61]			●	IID	-
FedRecover[4]	●		●	Non-IID	Fang
Wu et al.[64]	●			n/d	n/d
FedRecovery[77]	●			IID	-
MetaFul[59]	●			IID, Non-IID	Dirichlet
Deng et al.[9]	●			IID	-
Crab[24]	●			n/d	-
FedEraser[34]			●	n/d	n/d
FRU[75]	●	●	●	n/d	n/d
SIFU[15]	●		●	IID, Non-IID	Dirichlet
SecForget[36]		●		n/d	n/d
FFMU[6]		●		n/d	n/d
FedFilter[60]	●			Non-IID	-
UKRL[70]		●		IID, Non-IID	random
MoDe[80]	●	●		Non-IID	Dirichlet
FRAMU[50]		●	●	Non-IID	concept drift
VeriFi[16]	●			Non-IID	Dirichlet
Lin et al.[33]	●			n/d	
FC[46]	●			IID, Non-IID	n/d
Wang et al.[58]	●	●	●	IID, Non-IID	Fang
SecureCut[76]	●	●		n/d	n/d
FAST[20]	●			IID, Non-IID	random
ElBedoui et al.[12]		●		IID	-
FedME2[67]		●	●	n/d	n/d
Alam et al.[1]		●		IID	-
BFU[62]		●		n/d	n/d
FedHarmony[11]		●	●	Non-IID	covariate shift
2F2L[25]		●		IID	-
Liu et al.[38]		●		IID	-
FedLU[81]	●	●	●	Non-IID	unique
FedAF[31]		●		n/d	n/d
HDUS[73]			●	Non-IID	unique
EWC-SGA[65]		●		IID, Non-IID	unique
SFU[29]	●	●	●	IID, Non-IID	Dirichlet
Halimi et al.[21]		●		IID	-
QuickDrop[10]		●	●	IID, Non-IID	Dirichlet
forget-SVGD[17]		●		Non-IID	unique
Cforget-SVGD[18]		●		Non-IID	unique
KNOT[53]	●	●	●	Non-IID	Dirichlet
Lin et al.[32]			●	IID, Non-IID	random

Who Unlearns under What Data Distribution?

Unlearner	Global	Own local	All local	Raw data
Server	✓		✓	
Target client	✓	✓		✓
Remaining clients	✓	✓		

Who Unlearns?

- Available knowledge varies depending on who unlearns.
- What if a target request removal and leave?

Data Distribution?

- Only 54% considered Non-IID data settings.
- Non-IID simulation ≠ Real world data.

Ref.	Unlearner			Data Dist.	NIID sim.
	Server	Target	Remain		
RevFRF[37]	●			n/d	n/d
Exact-Fun[68]		●	●	Non-IID	random
FATS[56]	●			Non-IID	Dirichlet
Shao et al.[52]	●			Non-IID	unique
Wang et al.[61]			●	IID	-
FedRecover[4]	●		●	Non-IID	Fang
Wu et al.[64]	●			n/d	n/d
FedRecovery[77]	●			IID	-
MetaFul[59]	●			IID, Non-IID	Dirichlet
Deng et al.[9]	●			IID	-
Crab[24]	●			n/d	-
FedEraser[34]			●	n/d	n/d
FRU[75]	●	●	●	n/d	n/d
SIFU[15]	●		●	IID, Non-IID	Dirichlet
SecForget[36]		●		n/d	n/d
FFMU[6]		●		n/d	n/d
FedFilter[60]	●			Non-IID	-
UKRL[70]		●		IID, Non-IID	random
MoDe[80]	●	●		Non-IID	Dirichlet
FRAMU[50]		●	●	Non-IID	concept drift
VeriFi[16]	●			Non-IID	Dirichlet
Lin et al.[33]	●			n/d	
FC[46]	●			IID, Non-IID	n/d
Wang et al.[58]	●	●	●	IID, Non-IID	Fang
SecureCut[76]	●	●		n/d	n/d
FAST[20]	●			IID, Non-IID	random
ElBedoui et al.[12]		●		IID	-
FedME2[67]		●	●	n/d	n/d
Alam et al.[1]		●		IID	-
BFU[62]		●		n/d	n/d
FedHarmony[11]		●	●	Non-IID	covariate shift
2F2L[25]		●		IID	-
Liu et al.[38]		●		IID	-
FedLU[81]	●	●	●	Non-IID	unique
FedAF[31]		●		n/d	n/d
HDUS[73]			●	Non-IID	unique
EWC-SGA[65]		●		IID, Non-IID	unique
SFU[29]	●	●	●	IID, Non-IID	Dirichlet
Halimi et al.[21]		●		IID	-
QuickDrop[10]		●	●	IID, Non-IID	Dirichlet
forget-SVGD[17]		●		Non-IID	unique
Cforget-SVGD[18]		●		Non-IID	unique
KNOT[53]	●	●	●	Non-IID	Dirichlet
Lin et al.[32]			●	IID, Non-IID	random

On What Dataset?

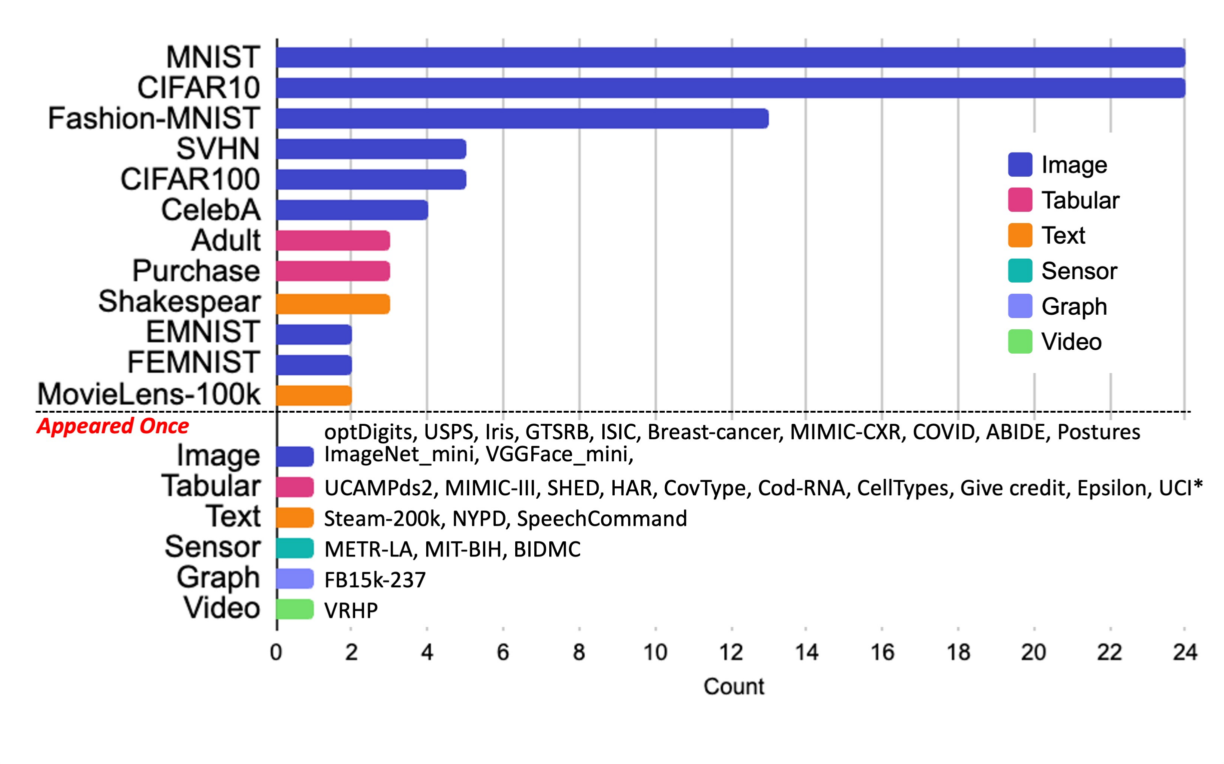


Table 4: Counts of data types used for experiments.

Data Type	Count	Modality	Count
Image	90	Uni-modal	123
Tabular	23	Multi-modal	2
Text	6	"Other" includes 3 sensors, 1 graph, 1 3D modeling, and 1 video dataset.	
Other	6		

*** The total count is 125.**

On What Dataset?

University of Massachusetts Amherst

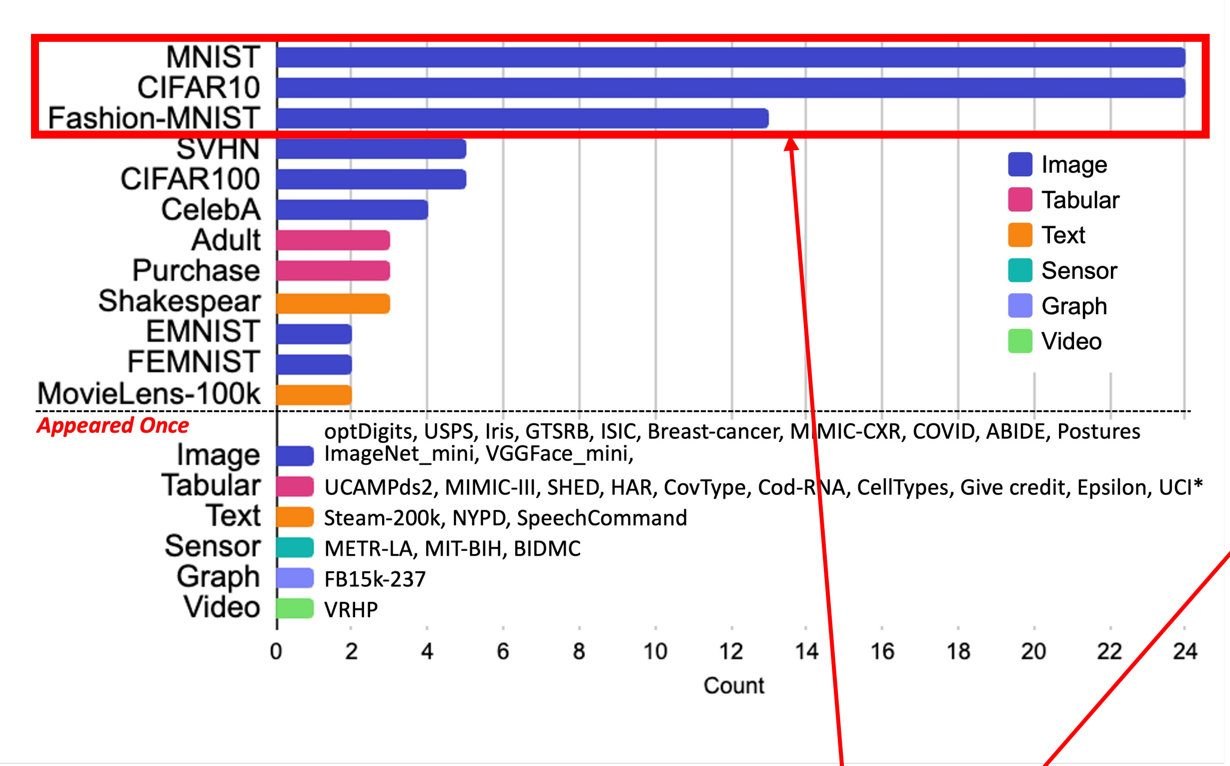


Table 4: Counts of data types used for experiments.

Data Type	Count	Modality	Count
Image	90	Uni-modal	123
Tabular	23	Multi-modal	2
Text	6	"Other" includes 3 sensors, 1 graph, 1 3D modeling, and 1 video dataset.	
Other	6		

*** The total count is 125.**

Mostly on (simple) **image** datasets for classification tasks.

Learning Configurations

Model architecture?

- Mostly simple CNNs
- Less use of pretrained models

Ref.	Data Type				Model Architecture	Aggregation Method
	im	ta	tx	ot		
RevFRF[37]	●	●			Random Forest	n/d
Exact-Fun[68]	●				3-, 4-layer CNN	FedAvg
FATS[56]	●		●		CNN, pretrained VGG16, LSTM	FedAvg
Shao et al.[52]	●				LeNet5	Weighted Avg
Wang et al.[61]				●	Linear model	FedAvg
FedRecover[4]	●	●			3-layer CNN, FCNN	FedAvg, Med, TrMean
Wu et al.[64]	●				2-layer CNN, VGG11, AlexNet	FedAvg
FedRecovery[77]	●				pre-trained CNN	FedAvg
MetaFul[59]	●			●	VGG16, LSTM	FedAvg
Deng et al.[9]	●	●			CNN	n/d
Crab[24]	●		●		n/d	FedAvg
FedEraser[34]	●	●			MLP, 4-layer CNN	FedAvg
FRU[75]			●		NCF, LightGCN	FedAvg
SIFU[15]	●				Regression model, CNN	FedAvg
FFMU[6]	●				CNN, LeNet, ResNet18	FedAvg
FedFilter[60]			●		4-layer CNN	Avg. base layers
UKRL[70]	●				DNN	FedAvg
MoDe[80]	●				ResNet	FedAvg
FRAMU[50]	●	●	●	●	n/d	FedAvg
VeriFi[16]	●		●		LeNet5, ResNet18, CNN, DenseNet121	FedAvg, Krum, Median
Lin et al.[33]	●				3-, 4-layer CNN	Weighted Avg
FC[46]	●	●	●		DC-KMeans	SCMA
Wang et al.[58]	●				ResNet, pre-trained VGG	FedAvg
SecureCut[76]		●			Gradient Boosted Decision Tree (GBDT)	n/d
FAST[20]	●				MLP, 2-layer CNN, VGG11, MobileNet	FedAvg
Elbedoui et al.[12]				●	3-layer CNN	FedAvg
FedME2[67]	●				MobileNetv3-large, ResNet50, RegNet-8gf	FedAvg
Alam et al.[1]	●				VGG11, ResNet18	FedAvg
BFU[62]	●				3-layer BNN, ResNet18	FedAvg
FedHarmony[11]	●				VGG-based CNN	FedEqual
2F2L[25]	●				3-layer CNN	FedAvg
Liu et al.[38]	●				3-layer CNN, AlexNet, ResNet	FedAvg
FedLU[81]				●	TransE, ComplEx, RotE	FedAvg
FedAF[31]	●				3-layer CNN, ResNet10	FedAvg
HDUS[73]	●				ResNet8, 18, 50, MobileNet-S, -M, -L	n/d
EWC-SGA[65]	●				n/d	FedAvg
SFU[29]	●				MLP, 3-layer CNN, ResNet18	n/d
Halimi et al.[21]	●				3-layer CNN	FedAvg
QuickDrop[10]	●				3-layer CNN	FedAvg
forget-SVGD[17]	●				1-layer BNN	n/d
Cforget-SVGD[18]	●				MLP	FedAvg
KNOT[53]	●	●	●		VGG16, LeNet5, MLP, GPT2	FedAvg, FedBuff
Lin et al.[32]	●		●		3-layer CNN, NanoGPT	FedAvg

Learning Configurations

Model architecture?
- Mostly simple CNNs
- Less use of pretrained models

Ref.	Data Type				Model Architecture	Aggregation Method
	im	ta	tx	ot		
RevFRF[37]	●	●			Random Forest	n/d
Exact-Fun[68]	●				3-, 4-layer CNN	FedAvg
FATS[56]	●		●		CNN, pretrained VGG16, LSTM	FedAvg
Shao et al.[52]	●				LeNet5	Weighted Avg
Wang et al.[61]				●	Linear model	FedAvg
FedRecover[4]	●	●			3-layer CNN, FCNN	FedAvg, Med, TrMean
Wu et al.[64]	●				2-layer CNN, VGG11, AlexNet	FedAvg
FedRecovery[77]	●				pre-trained CNN	FedAvg
MetaFul[59]	●			●	VGG16, LSTM	FedAvg
Deng et al.[9]	●	●			CNN	n/d
Crab[24]	●		●		n/d	FedAvg
FedEraser[34]	●	●			MLP, 4-layer CNN	FedAvg
FRU[75]			●		NCF, LightGCN	FedAvg
SIFU[15]	●				Regression model, CNN	FedAvg
FFMU[6]	●				CNN, LeNet, ResNet18	FedAvg
FedFilter[60]			●		4-layer CNN	Avg. base layers
UKRL[70]	●				DNN	FedAvg
MoDe[80]	●				ResNet	FedAvg
FRAMU[50]	●	●	●	●	n/d	FedAvg
VeriFi[16]	●		●		LeNet5, ResNet18, CNN, DenseNet121	FedAvg, Krum, Median
Lin et al.[33]	●				3-, 4-layer CNN	Weighted Avg
FC[46]	●	●	●		DC-KMeans	SCMA
Wang et al.[58]	●				ResNet, pre-trained VGG	FedAvg
SecureCut[76]		●			Gradient Boosted Decision Tree (GBDT)	n/d
FAST[20]	●				MLP, 2-layer CNN, VGG11, MobileNet	FedAvg
Elbedoui et al.[12]				●	3-layer CNN	FedAvg
FedME2[67]	●				MobileNetv3-large, ResNet50, RegNet-8gf	FedAvg
Alam et al.[1]	●				VGG11, ResNet18	FedAvg
BFU[62]	●				3-layer BNN, ResNet18	FedAvg
FedHarmony[11]	●				VGG-based CNN	FedEqual
2F2L[25]	●				3-layer CNN	FedAvg
Liu et al.[38]	●				3-layer CNN, AlexNet, ResNet	FedAvg
FedLU[81]				●	TransE, ComplEx, RotE	FedAvg
FedAF[31]	●				3-layer CNN, ResNet10	FedAvg
HDUS[73]	●				ResNet8, 18, 50, MobileNet-S, -M, -L	n/d
EWC-SGA[65]	●				n/d	FedAvg
SFU[29]	●				MLP, 3-layer CNN, ResNet18	n/d
Halimi et al.[21]	●				3-layer CNN	FedAvg
QuickDrop[10]	●				3-layer CNN	FedAvg
forget-SVGD[17]	●				1-layer BNN	n/d
Cforget-SVGD[18]	●				MLP	FedAvg
KNOT[53]	●	●	●		VGG16, LeNet5, MLP, GPT2	FedAvg, FedBuff
Lin et al.[32]	●		●		3-layer CNN, NanoGPT	FedAvg

Learning Configurations

Model architecture?

- Mostly simple CNNs
- Less use of pretrained models

Aggregation methods?

- Simple FedAvg (> 90% of works)
- Median, Trimmed Mean

Ref.	Data Type				Model Architecture	Aggregation Method
	im	ta	tx	ot		
RevFRF[37]	●	●			Random Forest	n/d
Exact-Fun[68]	●				3-, 4-layer CNN	FedAvg
FATS[56]	●		●		CNN, pretrained VGG16, LSTM	FedAvg
Shao et al.[52]	●				LeNet5	Weighted Avg
Wang et al.[61]				●	Linear model	FedAvg
FedRecover[4]	●	●			3-layer CNN, FCNN	FedAvg, Med, TrMean
Wu et al.[64]	●				2-layer CNN, VGG11, AlexNet	FedAvg
FedRecovery[77]	●				pre-trained CNN	FedAvg
MetaFul[59]	●			●	VGG16, LSTM	FedAvg
Deng et al.[9]	●	●			CNN	n/d
Crab[24]	●		●		n/d	FedAvg
FedEraser[34]	●	●			MLP, 4-layer CNN	FedAvg
FRU[75]			●		NCF, LightGCN	FedAvg
SIFU[15]	●				Regression model, CNN	FedAvg
FFMU[6]	●				CNN, LeNet, ResNet18	FedAvg
FedFilter[60]			●		4-layer CNN	Avg. base layers
UKRL[70]	●				DNN	FedAvg
MoDe[80]	●				ResNet	FedAvg
FRAMU[50]	●	●	●	●	n/d	FedAvg
VeriFi[16]	●		●		LeNet5, ResNet18, CNN, DenseNet121	FedAvg, Krum, Median
Lin et al.[33]	●				3-, 4-layer CNN	Weighted Avg
FC[46]	●	●	●		DC-KMeans	SCMA
Wang et al.[58]	●				ResNet, pre-trained VGG	FedAvg
SecureCut[76]		●			Gradient Boosted Decision Tree (GBDT)	n/d
FAST[20]	●				MLP, 2-layer CNN, VGG11, MobileNet	FedAvg
Elbedoui et al.[12]				●	3-layer CNN	FedAvg
FedME2[67]	●				MobileNetv3-large, ResNet50, RegNet-8gf	FedAvg
Alam et al.[1]	●				VGG11, ResNet18	FedAvg
BFU[62]	●				3-layer BNN, ResNet18	FedAvg
FedHarmony[11]	●				VGG-based CNN	FedEqual
2F2L[25]	●				3-layer CNN	FedAvg
Liu et al.[38]	●				3-layer CNN, AlexNet, ResNet	FedAvg
FedLU[81]				●	TransE, ComplEx, RotE	FedAvg
FedAF[31]	●				3-layer CNN, ResNet10	FedAvg
HDUS[73]	●				ResNet8, 18, 50, MobileNet-S, -M, -L	n/d
EWC-SGA[65]	●				n/d	FedAvg
SFU[29]	●				MLP, 3-layer CNN, ResNet18	n/d
Halimi et al.[21]	●				3-layer CNN	FedAvg
QuickDrop[10]	●				3-layer CNN	FedAvg
forget-SVGD[17]	●				1-layer BNN	n/d
Cforget-SVGD[18]	●				MLP	FedAvg
KNOT[53]	●	●	●		VGG16, LeNet5, MLP, GPT2	FedAvg, FedBuff
Lin et al.[32]	●		●		3-layer CNN, NanoGPT	FedAvg

Research Implications

- Mostly focused on
  efficacy, fidelity, efficiency
- Less considerations on
  security, guarantee, adaptivity, scalability

Ref.	Aggregation Method	Implication						
		efc	fid	efn	sec	gua	ada	sca
RevFRF[37]	n/d	●	●	●	●			
Exact-Fun[68]	FedAvg	●	●	●				
FATS[56]	FedAvg	●	●	●		●		
Shao et al.[52]	Weighted Avg	●		●		●		
Wang et al.[61]	FedAvg	●	●					
FedRecover[4]	FedAvg, Med, TrMean	●		●	●			
Wu et al.[64]	FedAvg	●	●					
FedRecovery[77]	FedAvg	●	●	●	●	●		
MetaFul[59]	FedAvg	●	●	●				
Deng et al.[9]	n/d	●	●	●	●			
Crab[24]	FedAvg	●	●	●				
FedEraser[34]	FedAvg	●	●	●				
FRU[75]	FedAvg	●	●	●			●	●
SIFU[15]	FedAvg	●	●	●	●	●		
FFMU[6]	FedAvg	●	●	●		●		
FedFilter[60]	Avg. base layers	●						
UKRL[70]	FedAvg	●		●	●			
MoDe[80]	FedAvg	●	●	●				
FRAMU[50]	FedAvg	●	●	●			●	
VeriFi[16]	FedAvg, Krum, Median	●	●	●	●			
Lin et al.[33]	Weighted Avg	●	●	●	●		●	
FC[46]	SCMA	●	●	●	●	●		
Wang et al.[58]	FedAvg	●	●	●				
SecureCut[76]	n/d	●	●		●			
FAST[20]	FedAvg	●	●	●	●			
Elbedoui et al.[12]	FedAvg	●						
FedME2[67]	FedAvg	●	●		●			
Alam et al.[1]	FedAvg	●	●					
BFU[62]	FedAvg	●	●	●				
FedHarmony[11]	FedEqual	●						
2F2L[25]	FedAvg	●	●					
Liu et al.[38]	FedAvg	●	●	●		●	●	
FedLU[81]	FedAvg	●	●	●				
FedAF[31]	FedAvg	●	●	●	●			
HDUS[73]	n/d	●					●	●
EWC-SGA[65]	FedAvg	●	●					
SFU[29]	n/d	●	●		●			
Halimi et al.[21]	FedAvg	●	●	●				
QuickDrop[10]	FedAvg	●	●	●				●
forget-SVGD[17]	n/d	●	●					
Cforget-SVGD[18]	FedAvg	●	●	●				
KNOT[53]	FedAvg, FedBuff	●	●	●			●	●
Lin et al.[32]	FedAvg	●	●	●		●	●	●

Evaluation Objectives and Metrics

University of Massachusetts Amherst

Table 7: A summary of Evaluation Metrics

Objective	Category	Metric
Efficacy	Performance	Accuracy on the target set Loss and errors on the target set MSE and MAE
	Parameter difference	L2 distance KLD Error rate (SAPE, ECE) Angular deviation 1st Wasserstein distance
	Indiscrimination quality	ASR, precision, and recall on BA ASR, precision, and recall on MIA Multi-task learning Influence function
Fidelity	Performance	Accuracy on test set Accuracy on remaining dataset Loss and errors on remaining set
Efficiency	Complexity	Time taken for unlearning Speed-up ratio Memory in MB

- Compare **retraining** and **unlearning**

No benchmark metric to assess different approaches

Evaluation Objectives and Metrics

University of Massachusetts Amherst

Table 7: A summary of Evaluation Metrics

Objective	Category	Metric
Efficacy	Performance	Accuracy on the target set Loss and errors on the target set MSE and MAE
	Parameter difference	L2 distance KLD Error rate (SAPE, ECE) Angular deviation 1st Wasserstein distance
	Indiscrimination quality	ASR, precision, and recall on BA ASR, precision, and recall on MIA Multi-task learning Influence function
Fidelity	Performance	Accuracy on test set Accuracy on remaining dataset Loss and errors on remaining set
Efficiency	Complexity	Time taken for unlearning Speed-up ratio Memory in MB

- Compare **retraining** and **unlearning**

No benchmark metric to assess different approaches

Simple BAs obscured impact of unlearning

Insights and Future Research Direction

University of Massachusetts Amherst

- Data are **heterogeneous**.
- Privacy-preserving unlearning is needed in **many domain**.
- **Advanced aggregation** methods could alleviate issues.
- FU introduces **additional** privacy **vulnerabilities**.
- **Benchmark evaluation** metrics enable method comparisons against a common standard.
- **Simple BA** impacts reduces by training round.

Questions & Answers

Full paper: https://arxiv.org/abs/2403.02437

University of Massachusetts Amherst