Rethinking Evaluation Metrics for Machine Unlearning

Yingdan Shi [1] **Sijia Liu** [2] **Ren Wang** [1]

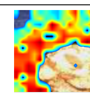
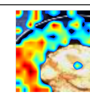
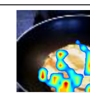
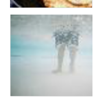
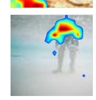
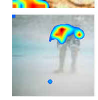
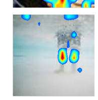
Abstract

Machine unlearning aims to eliminate the impact of specific data on a trained model. Although metrics like unlearning accuracy (UA) and membership inference attack (MIA) are commonly used to evaluate forgetting quality, they fall short in capturing the reliability of forgetting. In this work, we observe that even when data are misclassified according to UA and MIA, their ground truth labels can still remain within the predictive set from an uncertainty quantification perspective, revealing a fake unlearning issue. To better assess forgetting quality, we propose two novel metrics inspired by conformal prediction that offer a more faithful evaluation of forgetting reliability. Building upon these insights, we further introduce a conformal prediction-guided unlearning framework that integrates the Carlini & Wagner adversarial loss. This framework effectively encourages the exclusion of ground truth labels from the conformal prediction set. Extensive experiments on image classification tasks demonstrate the effectiveness of our proposed metrics. By incorporating a tailored loss term, our unlearning framework improves the UA of existing unlearning methods by an average of 6.6%.

1. Introduction

Machine unlearning is critical for ensuring data privacy, particularly under regulations like the General Data Protection Regulation (GDPR) (Bourtoule et al., 2021), which emphasize the right to erase personal data. It also serves as a tool to remove harmful biases and stereotypes embedded in models. Existing post hoc unlearning methods are broadly categorized into training-based (Graves et al., 2021; Tarun et al., 2023; Thudi et al., 2022; Warnecke et al., 2021) and training-free (Foster et al., 2024; Golatkar et al., 2021; 2020; Guo et al., 2019; Nguyen et al., 2020; Sekhari et al., 2021) approaches, depending on whether they require retraining during the unlearning process (Foster et al., 2024).

Table 1: Grad-CAM maps of one original model in CIFAR-10 with ResNet18 and two corresponding unlearning models. The **Prediction** row indicates whether the model correctly predicts the image's true label, while the **In Set** row represents whether the true label is included in the prediction set. Although the Finetune unlearning method, can misclassify the forget data, Grad-CAM can still highlight key features of the object under this model since the true label is included in the prediction set. In contrast, our unlearning method removes the true label from the set, with activation regions shifting significantly away from the object's key features. This confirms that the forgetting quality is better if the true label can be excluded from the prediction set.

Class Name	Forget Data	Original Model	Finetune Method	Our Method
Wok				
Swimming Trunks				
Prediction	–	✓	✗	✗
In Set	–	✓	✓	✗

To assess unlearning quality and model performance, several metrics have been proposed (Brophy & Lowd, 2021; Cao & Yang, 2015; Chen et al., 2021; Kashef, 2021; Shokri et al., 2017). However, widely used metrics such as unlearning accuracy (UA) and membership inference attack (MIA) are limited, as they mainly capture prediction correctness rather than the extent of forgetting. In particular, misclassification alone does not imply successful forgetting. To illustrate this, we apply conformal prediction (Papadopoulos et al., 2002; Lei & Wasserman, 2014) to unlearning models and observe that over 50% misclassified forget data instances still appear in the conformal prediction set. Grad-CAM visualizations (Selvaraju et al., 2017) in Table 1 confirm that although Finetune method misclassifies the forget data, the model still activates key features of the forgotten object.

Based on these insights, we introduce two new metrics that better quantify the uncertainty and robustness of unlearning. Furthermore, motivated by Carlini & Wagner (C&W) attack

[1]Illinois Institute of Technology [2]Michigan State University. Correspondence to: Ren Wang <rwang74@iit.edu>.

Published at ICML 2025 Workshop on Machine Unlearning for Generative AI.

(Carlini & Wagner, 2017), we propose a general conformal prediction-based framework to enhance training-based unlearning methods. Grad-CAM maps in Table 1 show that once the true label is excluded from the prediction set, the model's attention shifts significantly, indicating more effective forgetting.

2. Enhancing Metrics for Machine Unlearning via Conformal Prediction

2.1. Preliminaries and Notations

Machine Unlearning. We consider two scenarios, *random data* and *class-wise forgetting*. Let $\mathcal{D}_{train}$ be the training set for the original model $\boldsymbol{\theta}_o$, which is split into forget data $\mathcal{D}_f$ and retain data $\mathcal{D}_r = \mathcal{D}_{train} \backslash \mathcal{D}_f$. In random forgetting, $\mathcal{D}_{test}$ denotes the test set. In class-wise forgetting, $\mathcal{D}_{tf}$ and $\mathcal{D}_{tr}$ are test samples from the forget and retain classes, respectively. Let $\boldsymbol{\theta}_u$ denote the model after unlearning $\mathcal{D}_f$.

Conformal Prediction. Conformal prediction quantifies uncertainty by generating prediction sets with guaranteed coverage (Angelopoulos & Bates, 2021). There are four steps after unlearning:

1. *Calibration Data*. Prepare held-out data $\mathcal{D}_c$ disjoint from training and test sets.
2. *Non-conformity Score*. Calculate non-conformity score $S(\boldsymbol{x}, y_i) = 1 - p_i(\boldsymbol{x})$, where $p_i(\boldsymbol{x})$ is the probability for class y_i.
3. *Threshold*. Threshold $\hat{q} = \text{Quantile}_{1-\alpha}(S(\boldsymbol{x}, y_t))$ using only ground truth's non-conformity scores on $\mathcal{D}_c$ with miscoverage $\alpha \in [0, 1]$.
4. *Prediction Set*. Set $\mathbb{C}(\boldsymbol{x}) = \{y_i : S(\boldsymbol{x}, y_i) \le \hat{q}\}$.

2.2. Identifying Limitations in Existing Metrics

Inspired by uncertainty quantification, it is easy to recover the data that is identified as forgotten by the metrics UA and MIA. We propose a conformal prediction-based *recovery* method. If the ground truth of forget data falls within the conformal prediction set, we consider the recovery successful. Thus, **fake unlearning is defined as the scenario where the data identified as forgotten by existing metrics can be recovered by conformal prediction**.

In Table 2, we calculate the number of data points that are identified as truly forgotten by UA and MIA (marked as *mis-label*) and calculate how many of these points can still be recovered (marked as *in-set*) for retraining (RT), fine-tuning (FT) and random labeling (RL) methods. The **results of UA** reveal that even though the model misclassifies part of the forget data, on average 54.6% of these misclassified data instances are still recovered by conformal prediction. A similar phenomenon occurs on **results of MIA**. The *recover ratio* indicates that, although the MIA fails to identify an average of 18.33% of the forget data as training membership, conformal prediction can still recover 54.7% of these forget data within prediction sets.

Table 2: Mis-label count and in-set ratio of UA and MIA metrics for RT, FT and RL on **CIFAT-10** with **ResNet-18** under **10**$\%$ and **50**$\%$ **random data forgetting** scenarios.

Methods	10% Forgetting UA↑	Mis-label↑	In-set↓	Ratio↓	50% Forgetting MIA↓	Mis-label↑	In-set↓	Ratio↓
Mis-label and In-set Ratio of UA								
RT	8.62%	431	132	30.6%	10.98%	2,745	1,573	57.3%
FT	3.84%	192	112	58.3%	2.59%	647	431	66.6%
RL	7.55%	380	173	45.5%	10.48%	2,625	1,795	68.4%
Mis-label and In-set Ratio of MIA								
RT	86.92%	654	209	32.0%	82.79%	4,303	1,391	32.3%
FT	92.00%	400	216	54.0%	92.92%	1,769	813	46.0%
RL	74.21%	1,289	1,011	78.4%	61.15%	9,713	8,295	85.4%

Overall, the high *recover ratio* across both UA and MIA in Tables 2 highlights that these methods are not fully effective at eliminating traces of the forget data from the uncertainty quantification perspective and can still be recovered by conformal prediction. This encloses that **the fake unlearning issue arises when the ground truth label of misclassified data falls within the conformal prediction set**.

2.3. Designing Metrics via Conformal Prediction

Based on the limitation of UA and MIA metrics shown in Section 2.2, we propose two novel metrics that can capture fake unlearning.

Conformal Ratio (CR). To overcome the fake unlearning inherent in UA, we introduce a novel metric, CR, which incorporates both coverage and set size in conformal prediction to provide a more comprehensive evaluation. Before defining CR, we introduce Coverage and Set Size.

Given a dataset $\mathcal{D}$, the definition of **Coverage** is as follows:

$$\text{Coverage} := \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{x}, y_t) \in \mathcal{D}} \mathbb{I}(y_t \in \mathbb{C}(\boldsymbol{x})), \tag{1}$$

where y_t is the true label of data point $\boldsymbol{x}$. Indicator function $\mathbb{I}(\cdot)$ returns 1 if the enclosed condition is true and 0 otherwise. Coverage reflects the probability that the true label falls within the prediction set $\mathbb{C}(\boldsymbol{x})$. For $\mathcal{D} = \mathcal{D}_f$, high coverage indicates that the model retains significant information about forget data, suggesting fake unlearning.

Given a dataset $\mathcal{D}$, **Set Size** is defined as follows:

$$\text{Set Size} := \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{x}, y_t) \in \mathcal{D}} |\mathbb{C}(\boldsymbol{x})|, \tag{2}$$

where $|\mathbb{C}(\boldsymbol{x})|$ represents the prediction set size of a specific instance. When $y_t \in \mathbb{C}(\boldsymbol{x})$, a small set size indicates that fewer non-ground truth classes are included in the prediction set, reflecting stronger fake unlearning.

Based on Coverage and Set Size, we introduce the definition of **CR** for a dataset $\mathcal{D}$ as follows:

$$\text{CR} := \frac{\text{Coverage}}{\text{Set Size}} = \frac{\sum_{(\boldsymbol{x},y_t)\in\mathcal{D}} \mathbb{I}(y_t \in \mathbb{C}(\boldsymbol{x}))}{\sum_{(\boldsymbol{x},y_t)\in\mathcal{D}} |\mathbb{C}(\boldsymbol{x})|}. \quad (3)$$

CR balances the information captured by Coverage and Set Size. A lower CR value implies stronger unlearning on a given dataset. CR is inspired by conformal prediction which is proposed to assess the model's behavior on new and unseen data, not on the training data. Thus, we emphasize that CR only measures forget data $\mathcal{D}_f$ and test data $\mathcal{D}_{test}$.

MIA Conformal Ratio (MIACR). MIACR is proposed to address the limitation of MIA metric. Among three potential conformal prediction sets $\{0\}$, $\{1\}$ and $\{0, 1\}$, only set $\{0\}$ is ideal case for MIA, because the presence of '1' represents the data point can still be recognised as a training member. Therefore, we introduce a new metric **MIACR** as:

$$\text{MIACR} := \frac{1}{|\mathcal{D}_f|} \sum_{(\boldsymbol{x},y_t)\in\mathcal{D}_f} \mathbb{I}(\mathbb{C}(\boldsymbol{x}) = \{0\}), \quad (4)$$

where $\mathbb{C}(\boldsymbol{x}) = \{0\}$ denotes prediction set is $\{0\}$. A higher MIACR score indicates a stronger forgetting. While MIA consider a data point forgotten once the logit for label '0' exceeds that for label '1'. MIACR applies a stricter rule by also requiring that label '1' be absent from the conformal prediction set, giving a more rigorous check of membership status and forgetting quality.

Evaluation Criteria. Similar to existing unlearning metrics, we consider two different criteria[1] to measure performance with our metrics: ❶ **Gap to RT Criterion**: A lower gap to the RT method is better for both CR and MIACR metrics. ❷ **Limit-Based Criterion**: A lower CR value on $\mathcal{D}_f$ indicates stronger forgetting performance, while a higher MIACR value reflects better unlearning effectiveness.

3. Enhancing Machine Unlearning via Conformal Prediction

Existing unlearning methods are often optimized for loss functions that do not ensure the exclusion of true labels from the conformal prediction set. This motivates us to explore advanced unlearning techniques to enhance unlearning.

We propose a <u>c</u>onformal <u>p</u>rediction-based <u>u</u>nlearning framework (CPU), which encourages pushing the ground truth label's non-conformity score beyond the CP threshold $\hat{q}$.

[1]The appropriate evaluation criteria vary across unlearning applications (Kurmanji et al., 2024): criterion ❶ is relevant for user privacy, while criterion ❷ focuses on bias removal.

We first adapt the original C&W loss for unlearning:

$$\mathcal{L}_{\text{cw}}(\boldsymbol{x}, y_t) = \max\{p_t(\boldsymbol{x}) - \max_{i\neq t} p_i(\boldsymbol{x}),\ -\Delta\}, \quad (5)$$

where $\max\{\cdot\}$ is a maximum operator that selects the largest value from the set. $p_t(\boldsymbol{x})$ is the probability for the true label, and Δ controls the separation margin. We denote $\max_{i\neq t}\{p_i(\boldsymbol{x})\}$ as the highest probability value of the non-ground truth classes. This loss tries to decrease the probability of true class y_t and further increase that of the class y_i. While $\mathcal{L}_{\text{cw}}$ reduces $p_t(\boldsymbol{x})$, it doesn't guarantee $y_t \notin \mathbb{C}(\boldsymbol{x})$. To bridge this, we refine the loss based on conformal prediction insight and define the loss as:

$$\mathcal{L}_{\text{unlearn}}(\boldsymbol{x}, y_t) = \max\{\bar{q} - S(\boldsymbol{x}, y_t),\ -\Delta\}, \quad (6)$$

where $S(\boldsymbol{x}, y_t)$ is non-conformity score of true label y_t. This encourages $S(\boldsymbol{x}, y_t) - \bar{q} \geq \Delta$. It helps to increase non-conformity score $S(\boldsymbol{x}, y_t)$ of true label y_t to surpass the threshold $\bar{q}$ and push the true label out of conformal prediction set. To preserve the original unlearning strategy, our final loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{original}} + \lambda \cdot \mathcal{L}_{\text{unlearn}}, \quad (7)$$

where λ balances forgetting and original objectives.

4. Experiment

4.1. Experimental Setting

We report experiments on CIFAR-10 (Krizhevsky, 2009) with ResNet-18 (He et al., 2016) and Tiny ImageNet (Le & Yang, 2015) with ViT (Dosovitskiy et al., 2021). We employ 9 different unlearning methods, **RT**, **FT** (Warnecke et al., 2021), **RL** (Graves et al., 2021), **Gradient Ascent (GA)** (Thudi et al., 2022), **Teacher** (Tarun et al., 2023), **SSD** (Foster et al., 2024), **NegGrad+** (Kurmanji et al., 2024), **Salun** (Fan et al., 2023) and **SFRon** (Huang et al., 2025).

4.2. Experimental Results

CR Metric. We present the CR and MIACR results on CIFAR-10 in Table 3. See additional results on other settings, datasets and architectures in Appendix B. For CR metric, according to evaluation criterion ❶, the top 4 methods under the UA metric are *NegGrad+*, *RL*, *SFRon*, and *Salun*, as their unlearning accuracy is closest to the RT method. However, this ranking shifts slightly under the CR metric, where the top 4 become *Salun*, *NegGrad+*, *SFRon*, and *RL*. CR metric identifies that RL faces a fake unlearning situation and performs poorly on our metric CR. This observation suggests that methods excelling in traditional UA metric may not perform well under the CR metric. **The underlying rationale behind this is that the CR metric takes into account the possibility that the true labels of**

Table 3: Unlearning performance of 9 unlearning methods on **CIFAR-10** with **ResNet-18** in 10% **random data forgetting**. The sign $\uparrow$ represents the greater is better, while $\downarrow$ denotes ideally small. The results are average values from 3 independent trials and the standard deviation values are reported in Appendix B. The performance gap relative to the RT method is represented in (•) in results tables. It shows the unlearning methods that excel under UA metric do not necessarily perform well under the CR metric.

Methods	Existing Metrics				Coverage		Set Size		CR		MIACR
	UA $\uparrow$	RA $\uparrow$	TA $\uparrow$	MIA $\downarrow$	$\mathcal{D}_f \downarrow$	$\mathcal{D}_{test} \uparrow$	$\mathcal{D}_f \uparrow$	$\mathcal{D}_{test} \downarrow$	$\mathcal{D}_f \downarrow$	$\mathcal{D}_{test} \uparrow$	$\mathcal{D}_f \uparrow$
RT	8.6%(0.0)	99.7%(0.0)	91.8%(0.0)	86.92(0.000)	0.941(0.000)	0.944(0.000)	1.089(0.000)	1.074(0.000)	0.864(0.000)	0.879(0.000)	0.091(0.000)
FT	3.8%(4.8)	98.1%(1.6)	91.6%(0.2)	92.00(5.08)	0.994(0.053)	0.951(0.007)	1.008(0.081)	1.026(0.048)	0.986(0.122)	0.927(0.048)	0.091(0.000)
RL	7.6%(1.0)	97.4%(2.3)	90.6%(1.2)	74.21(12.71)	0.970(0.029)	0.949(0.005)	1.242(0.153)	1.197(0.123)	0.788(0.076)	0.796(0.083)	0.083(0.008)
GA	0.6%(8.0)	99.5%(0.2)	94.1%(2.3)	98.80(11.88)	0.994(0.053)	0.945(0.001)	1.002(0.087)	1.009(0.065)	0.994(0.130)	0.936(0.057)	0.012(0.079)
Teacher	0.8%(7.8)	99.4%(0.3)	93.5%(1.7)	87.24(0.32)	0.991(0.050)	0.941(0.003)	1.003(0.086)	1.021(0.053)	0.993(0.129)	0.922(0.043)	0.013(0.078)
SSD	0.5%(8.1)	99.5%(0.2)	94.2%(2.4)	98.78(11.86)	0.996(0.055)	0.945(0.001)	0.999(0.090)	1.008(0.066)	0.994(0.130)	0.936(0.057)	0.011(0.080)
NegGrad+	8.7%(0.1)	98.8%(0.9)	92.2%(0.4)	90.30(3.38)	0.934(0.007)	0.948(0.004)	1.068(0.021)	1.086(0.012)	0.875(0.011)	0.873(0.006)	0.076(0.015)
Salun	3.7%(4.9)	98.9%(0.8)	91.8%(0.0))	57.58(29.34)	0.987(0.046)	0.950(0.006)	1.132(0.043)	1.143(0.069)	0.872(0.008)	0.832(0.047)	0.055(0.036)
SFRon	4.8%(3.8)	97.4%(2.3)	91.4%(0.4)	91.55(4.63)	0.977(0.036)	0.953(0.009)	1.100(0.011)	1.143(0.069)	0.889(0.025)	0.834(0.045)	0.017(0.074)

Table 4: Performance of our framework CPU. We show the performance on **CIFAR-10** with **ResNet-18** and **Tiny ImageNet** with **ViT** in 10% **random data forgetting**. $\lambda = 0$ represents the baseline without our framework applied. It shows our framework significantly improves the forgetting quality, not only across our metric but also existing metric UA, while preserving stable predictive performance.

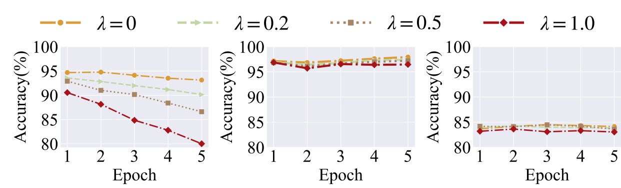
Methods	$\boldsymbol{\lambda = 0}$					$\boldsymbol{\lambda = 0.2}$					$\boldsymbol{\lambda = 0.5}$				
	UA $\uparrow$	RA $\uparrow$	TA $\uparrow$	$\text{CR}_{\mathcal{D}_f} \downarrow$	$\text{CR}_{\mathcal{D}_{test}} \uparrow$	UA $\uparrow$	RA $\uparrow$	TA $\uparrow$	$\text{CR}_{\mathcal{D}_f} \downarrow$	$\text{CR}_{\mathcal{D}_{test}} \uparrow$	UA $\uparrow$	RA $\uparrow$	TA $\uparrow$	$\text{CR}_{\mathcal{D}_f} \downarrow$	$\text{CR}_{\mathcal{D}_{test}} \uparrow$
CIFAR-10 with ResNet-18															
CPU-RT	8.6%(0.0)	99.7%(0.0)	91.8%(0.0)	0.864(0.000)	0.879(0.000)	10.8%(2.2)	98.3%(1.4)	91.0%(0.8)	0.788(0.076)	0.824(0.055)	14.0%(5.4)	97.8%(1.9)	90.4%(0.4)	0.763(0.101)	0.825(0.054)
CPU-FT	3.8%(4.8)	98.1%(1.6)	91.6%(0.2)	0.986(0.122)	0.927(0.048)	6.8%(1.8)	97.0%(2.7)	90.8%(1.0)	0.844(0.020)	0.829(0.050)	7.9%(0.7)	96.9%(2.8)	90.9%(0.9)	0.853(0.011)	0.843(0.036)
CPU-RL	7.6%(1.0)	97.4%(2.3)	90.6%(1.2)	0.788(0.076)	0.796(0.083)	9.7%(1.1)	96.6%(3.1)	89.4%(2.4)	0.709(0.155)	0.736(0.143)	9.9%(1.3)	96.9%(2.8)	89.7%(2.1)	0.708(0.156)	0.731(0.148)
Tiny ImageNet with ViT															
CPU-RT	14.7%(0.0)	98.8%(0.0)	86.0%(0.0)	0.503(0.000)	0.516(0.000)	19.3%(4.6)	98.8%(0.0)	86.0%(0.0)	0.458(0.045)	0.516(0.000)	26.4%(11.7)	98.7%(0.1)	85.8%(0.2)	0.396(0.107)	0.489(0.027)
CPU-FT	6.9%(7.8)	97.9%(0.9)	84.1%(1.9)	0.466(0.037)	0.389(0.127)	9.8%(4.9)	97.4%(1.4)	83.6%(2.4)	0.441(0.062)	0.399(0.117)	13.6%(0.9)	97.2%(1.6)	83.6%(2.4)	0.413(0.090)	0.401(0.115)
CPU-RL	26.9%(12.2)	96.0%(2.8)	81.4%(4.6)	0.054(0.449)	0.111(0.405)	31.8%(17.1)	95.3%(17.9)	80.9%(5.1)	0.051(0.452)	0.111(0.405)	36.2%(21.5)	95.3%(3.5)	80.4%(5.6)	0.051(0.452)	0.121(0.395)

some misclassified forget data points may still remain within the prediction set. This observation aligns with the insights we discussed in Section 2.2 regarding the fake unlearning issue of UA metric.

Regarding evaluation criterion ❷, a similar pattern is observed as with criterion ❶. Under the UA metric, the top 4 methods are *NegGrad+*, *RT*, *RL* and *SFRon*. However, under the CR metric, the top 4 shift to *RL*, *RT*, *Salun* and *NegGrad*. This indicates that some unlearning methods, such as NegGrad+ show weak forgetting quality when viewed from the fake unlearning perspective. This also highlights that the CR captures critical scenarios overlooked by UA, specifically the potential retention of true labels within prediction sets for the forget data points.

MIACR Metric. Under evaluation criterion ❶, the unlearning methods considered superior under MIA metric, such as Teacher and SFRon, fail to maintain their exceptional performance when evaluated using the MIACR criterion. Similarly, under evaluation criterion ❷, Salun, which is considered optimal according to MIA metrics, does not demonstrate the best performance under MIACR evaluation. While MIA fails to accurately predict approximately 50% of the forget data in Salun, MIACR can still identify membership within the prediction set.

Performance of Our Unlearning Framework In this experiment, we apply the RT, FT, and RL methods to our framework, i.e., CPU-RT, CPU-FT, CPU-RL. Table 4 presents the results for CIFAR-10 with ResNet-18 and Tiny ImageNet with ViT in 10% random data forgetting scenario. We vary λ in the range $[0, 0.2, 0.5]$, where $\lambda = 0$ represents



(a) $\mathcal{D}_f$ (b) $\mathcal{D}_r$ (c) $\mathcal{D}_{test}$

Figure 1: CPU-FT performance under different λ values across each epoch on Tiny ImageNet. As λ increases, accuracy on $\mathcal{D}_f$ drops significantly, while retain and test accuracy remain stable.

the baseline without our framework applied. The experimental results demonstrate a significant improvement in UA and $\text{CR}_{\mathcal{D}_f}$ metrics across all methods, reflecting improved forgetting quality as λ increases. Notably, the RA, TA, and $\text{CR}_{\mathcal{D}_{test}}$ values remain relatively stable, indicating that the substantial improvement in forgetting quality does not compromise the model's predictive performance. We demonstrate the CPU-FT results across each epoch on Tiny ImageNet in Figure 1, which also shows the same trend.

5. Conclusion

Motivated by conformal prediction, we introduce new metrics, CR and MIACR, to enhance the evaluation and reliability of machine unlearning. In addiction, our unlearning framework, which incorporates the adapted C&W loss with conformal prediction, improves unlearning effectiveness. Together, we provide a more rigorous foundation for privacy-preserving machine learning.

References

Angelopoulos, A. N. and Bates, S. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv preprint arXiv:2107.07511, 2021.

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., Lie, D., and Papernot, N. Machine unlearning. In 2021 IEEE Symposium on Security and Privacy (SP), pp. 141–159. IEEE, 2021.

Brophy, J. and Lowd, D. Machine unlearning for random forests. In International Conference on Machine Learning, pp. 1092–1104. PMLR, 2021.

Cao, Y. and Yang, J. Towards making systems forget with machine unlearning. In 2015 IEEE symposium on security and privacy, pp. 463–480. IEEE, 2015.

Carlini, N. and Wagner, D. Towards evaluating the robustness of neural networks. In 2017 ieee symposium on security and privacy (sp), pp. 39–57. Ieee, 2017.

Chen, M., Zhang, Z., Wang, T., Backes, M., Humbert, M., and Zhang, Y. When machine unlearning jeopardizes privacy. In Proceedings of the 2021 ACM SIGSAC conference on computer and communications security, pp. 896–911, 2021.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. URL https://arxiv.org/abs/2010.11929.

Fan, C., Liu, J., Zhang, Y., Wong, E., Wei, D., and Liu, S. Salun: Empowering machine unlearning via gradient-based weight saliency in both image classification and generation. arXiv preprint arXiv:2310.12508, 2023.

Foster, J., Schoepf, S., and Brintrup, A. Fast machine unlearning without retraining through selective synaptic dampening. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, pp. 12043–12051, 2024.

Golatkar, A., Achille, A., and Soatto, S. Forgetting outside the box: Scrubbing deep networks of information accessible from input-output observations. In European Conference on Computer Vision, pp. 383–398, 2020.

Golatkar, A., Achille, A., Ravichandran, A., Polito, M., and Soatto, S. Mixed-privacy forgetting in deep networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 792–801, 2021.

Graves, L., Nagisetty, V., and Ganesh, V. Amnesiac machine learning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pp. 11516–11524, 2021.

Guo, C., Goldstein, T., Hannun, A., and Van Der Maaten, L. Certified data removal from machine learning models. arXiv preprint arXiv:1911.03030, 2019.

He, K., Zhang, X., Ren, S., and Sun, J. Identity mappings in deep residual networks. In Leibe, B., Matas, J., Sebe, N., and Welling, M. (eds.), Computer Vision – ECCV 2016, pp. 630–645, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46493-0.

Huang, Z., Cheng, X., Zheng, J., Wang, H., He, Z., Li, T., and Huang, X. Unified gradient-based machine unlearning with remain geometry enhancement. Advances in Neural Information Processing Systems, 37:26377–26414, 2025.

Kashef, R. A boosted svm classifier trained by incremental learning and decremental unlearning approach. Expert Systems with Applications, 167:114154, 2021.

Krizhevsky, A. Learning multiple layers of features from tiny images. 2009. URL https://api.semanticscholar.org/CorpusID:18268744.

Kurmanji, M., Triantafillou, E., and Triantafillou, P. Machine unlearning in learned databases: An experimental analysis. Proc. ACM Manag. Data, 2(1), March 2024. doi: 10.1145/3639304. URL https://doi.org/10.1145/3639304.

Le, Y. and Yang, X. Tiny imagenet visual recognition challenge. CS 231N, 7(7):3, 2015.

Lei, J. and Wasserman, L. Distribution-free prediction bands for non-parametric regression. Journal of the Royal Statistical Society Series B: Statistical Methodology, 76 (1):71–96, 2014.

Nguyen, Q. P., Low, B. K. H., and Jaillet, P. Variational bayesian unlearning. Advances in Neural Information Processing Systems, 33:16025–16036, 2020.

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. Inductive confidence machines for regression. In Machine learning: ECML 2002: 13th European conference on machine learning Helsinki, Finland, August 19–23, 2002 proceedings 13, pp. 345–356. Springer, 2002.

Romano, Y., Sesia, M., and Candes, E. Classification with valid and adaptive coverage. Advances in Neural Information Processing Systems, 33:3581–3591, 2020.

Sekhari, A., Acharya, J., Kamath, G., and Suresh, A. T. Remember what you want to forget: Algorithms for machine unlearning. Advances in Neural Information Processing Systems, 34:18075–18086, 2021.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision, pp. 618–626, 2017.

Shokri, R., Stronati, M., Song, C., and Shmatikov, V. Membership inference attacks against machine learning models. In 2017 IEEE symposium on security and privacy (SP), pp. 3–18. IEEE, 2017.

Tarun, A. K., Chundawat, V. S., Mandal, M., and Kankanhalli, M. Deep regression unlearning. In International Conference on Machine Learning, pp. 33921–33939. PMLR, 2023.

Thudi, A., Deza, G., Chandrasekaran, V., and Papernot, N. Unrolling sgd: Understanding factors influencing machine unlearning. In 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P), pp. 303–319. IEEE, 2022.

Warnecke, A., Pirch, L., Wressnegger, C., and Rieck, K. Machine unlearning of features and labels. arXiv preprint arXiv:2108.11577, 2021.

A. Experimental Details

Setting Details. For CIFAR-10 with ResNet-18 architecture, we train the original model from scratch for 200 epochs using SGD with a Cosine Annealing learning rate schedule, starting from an initial learning rate of 0.1. We set the momentum to 0.9 and a batch size of 64. The RT model adopts the same training configuration. Other models are trained for the following durations: FT for 20 epochs, RL for 10 epochs, SalUn for 10 epochs, GA for 1 epoch (to avoid over-forgetting and significant RA degradation), NegGrad+ for 10 epochs (reduced to 2 epochs in class-wise scenarios), and SFRon for 10 epochs. All other hyperparameters match those of the original model.

For the ViT architecture, we initialize the original model by training a pretrained ViT model for 15 epochs on Tiny ImageNet. We start with a learning rate of 0.001, while other training parameters match those used for ResNet-18. We use SGD and set the momentum to 0.9 and a batch size of 64. The RT model follows the same training procedure as the original model. Other models are trained for the following durations: FT for 5 epochs, RL for 5 epochs, SalUn for 5 epochs, GA for 1 epoch, NegGrad+ for 5 epochs, and SFRon for 5 epochs. All other hyperparameters are consistent with the original model's training.

For CIFAR-10/Tiny ImageNet, we randomly select 200/50 data points per class (2000/10000 data points in total) as calibration data $\mathcal{D}_c$ and $\mathcal{D}'_c$, respectively. The calibration data $\mathcal{D}_c$ does not participate in the model training or unlearning processes and is only used for calibrating the threshold $\hat{q}$, while $\mathcal{D}'_c$ is used in the process of our unlearning framework to generate $\bar{q}$. All experiments are conducted on 1 Tesla V100-SXM2 GPU card with 32GB memory in a single node.

Confidence Level $1-\alpha$. A smaller miscoverage rate α, i.e., a higher confidence level $1-\alpha$, guarantees more reliable coverage, while a higher value of α results in a lower confidence level. In the conformal prediction related works (Angelopoulos & Bates, 2021; Papadopoulos et al., 2002; Romano et al., 2020), $\boldsymbol{\alpha = 0.05}$ has become the standard in most cases. This value can be rooted in its widespread adoption in statistical hypothesis testing, where it balances the trade-off between false positives and practical usability. Following previous work, we set α to 0.05, but we still report results for a higher range of α values (0.10, 0.15, and 0.20) in Appendix B to account for scenarios where a more relaxed confidence level may be needed. Unless otherwise stated, all subsequent analyzes use the recommended default of $\alpha = 0.05$.

Calibration Set Size. As for calibration data, a portion of the validation data is set aside as calibration data, ensuring it remains independent from both the training and test data. A key requirement in selecting the calibration set size is to avoid producing an abnormal threshold $\hat{q}$, as a small sample size may introduce outliers and lead to unstable coverage estimates. Therefore, it is essential to ensure a sufficiently calibration set to obtain reliable and stable estimates. Figure 2 illustrates the stability of $\hat{q}$ across varying calibration set sizes. The results are smoothed using B-spline. We implement them on CIFAR-10 with ResNet-18 in 10% and 50% random data forgetting scenarios. The results show that for different settings using ResNet-18 on CIFAR-10, after the calibration set size is larger than 1000, abnormal $\hat{q}$ values do not occur anymore, and a stable threshold $\hat{q}$ can be obtained. Similarly, we analyze the calibration set size of the class-wise forgetting scenario and find that fewer calibration data points are required compared to random data forgetting. This is because the targeted removal of specific classes reduces the complexity of the distribution, unlike the broader variability introduced by random data removal.

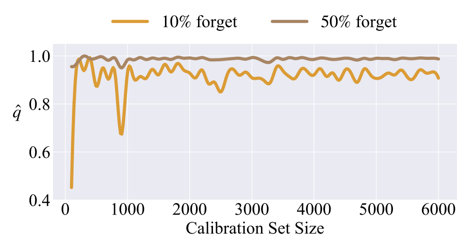


Figure 2: The stability of $\hat{q}$ in different calibration set sizes. When the calibration set size is greater than 2000, the fluctuations of $\hat{q}$ remain within a stable range.

B. Evaluating MU methods

B.1. Mis-label Number and In-set Ratios

Conformal prediction is applied to UA and MIA predictions to determine the number of misclassified data points (mis-label) and the number of these points that fall within the conformal prediction set (in-set) across 9 unlearning methods. We evaluate both the UA and MIA metrics by counting the misclassified data points and calculating how many of them are included in the conformal prediction set. The detailed results are presented in Table 5.

Table 5: Mis-label number and in-set ratios of UA and MIA metrics for 9 unlearning methods.

Methods	10% Forgetting Mis-label ↑	In-set ↓	Ratio ↓	50% Forgetting Mis-label ↑	In-set ↓	Ratio ↓
Mis-label and In-set Ratio of UA						
RT	431	132	30.6%	2,745	1,573	57.3%
FT	192	112	58.3%	647	431	66.6%
RL	380	173	45.5%	2,625	1,795	68.4%
GA	30	2	6.7%	150	9	6.0%
Teacher	40	4	10%	400	37	9.3%
FF	2,995	2,751	91.9%	15,083	14,061	93.2%
SSD	25	2	8.0%	116	9	7.8%
NegGrad+	435	115	26.4%	711	249	35.5%
Salun	185	117	63.2%	1065	695	65.3%
SFRon	240	125	52.1%	1000	610	61.0%
Mis-label and In-set Ratio of MIA						
RT	654	209	32.0%	4,303	1,391	32.3%
FT	400	216	54.0%	1,769	813	46.0%
RL	1,289	1,011	78.4%	9,713	8,295	85.4%
GA	60	10	16.7%	284	31	10.9%
Teacher	638	586	91.8%	1,689	895	53.0%
FF	1,424	1,424	100%	5,996	4,850	80.9%
SSD	61	11	18.0%	282	24	8.5%
NegGrad+	486	106	21.8%	1,545	415	26.9%
Salun	2,121	1,848	87.1%	10,221	9,121	89.2%
SFRon	423	121	28.6%	1,871	433	23.1%

B.2. Measuring Forgetting under Distribution Shifts

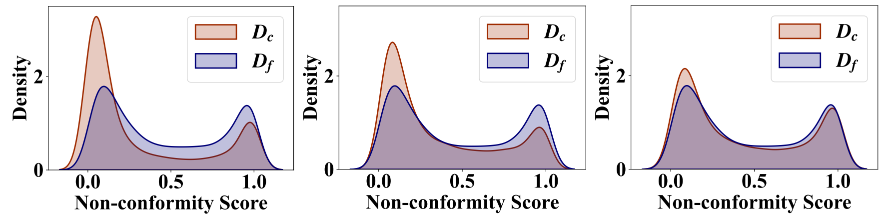
RL and Salun are unlearning methods employ label corruption in their unlearning strategy which can cause distribution shifts. Here, we introduce how to better measure forgetting under these circumstances. Figure 3(a) shows the non-conformity score distribution of calibration data $\mathcal{D}_c$ and forget data $\mathcal{D}_f$ in unlearning model θ_u obtained by RL method in Tiny ImageNet with ViT. It looks like there is a significant discrepancy between the distribution of the forget data and the calibration data.

To align the distribution of $\mathcal{D}_c$ with that of $\mathcal{D}_f$ and minimize the differences between them, we design a shadow model. To make the explanation clearer and more intuitive, we take RL as an example. In the RL unlearning method, the forget data is assigned random labels. Therefore, we apply the same random labeling process to the calibration data and train a shadow model accordingly. We designed two methods:

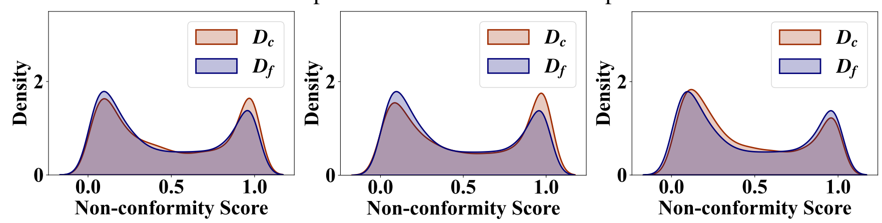
1. **Shadow model**. A shadow model replicates the behavior of forget data $\mathcal{D}_f$ throughout the unlearning process. A shadow model is a two-step approach: (1) it firstly trains a shadow original model θ'_o using train data $\mathcal{D}_{train}$ and clean calibration data $\mathcal{D}_c$ with the same epoch number as the original model θ_o; (2) subsequently, we finetune the θ'_o using the random labeled calibration data.

2. **Semi-shadow model**. The semi-shadow model only adopts the second step in the shadow model. It finetunes the original model θ_o with random-labeled calibration data.

The results are presented in Figure 3, where (b)-(e) present the results of the semi-shadow model with different epochs and (f) illustrates the shadow model's result. Under the semi-shadow model, as the number of epochs increases, the distribution of calibration data gradually moves to the right until it becomes consistent with the distribution of forget data. It also shows that the shadow model demonstrates the best ability to handle distribution shifts compared to the semi-shadow model. However, this comes at the cost of higher computational overhead. Overall, the semi-shadow model offers a balanced trade-off between handling distribution shifts effectively and maintaining lower computational costs.



(a) Without adjusting. (b) Semi-shadow with 1 epoch. (c) Semi-shadow with 3 epoch.



(d) Semi-shadow with 4 epoch. (e) Semi-shadow with 5 epoch. (f) Shadow with 5 epoch.

Figure 3: Distribution shifting processing with different strategies. The distribution of calibration data gradually converges with that of forget data.

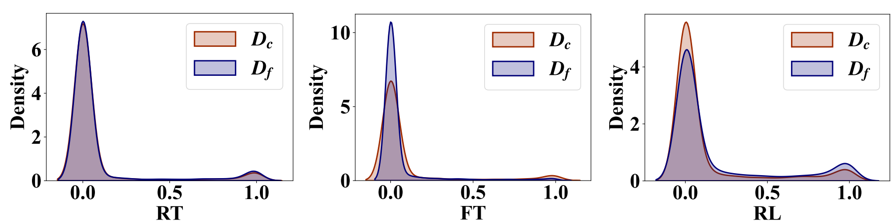


Figure 4: Non-conformity density of calibration data $\mathcal{D}_c$ and forget data $\mathcal{D}_f$ **without our unlearning framework** in CIFAR-10 with ResNet-18 under 10% random data forgetting scenario.

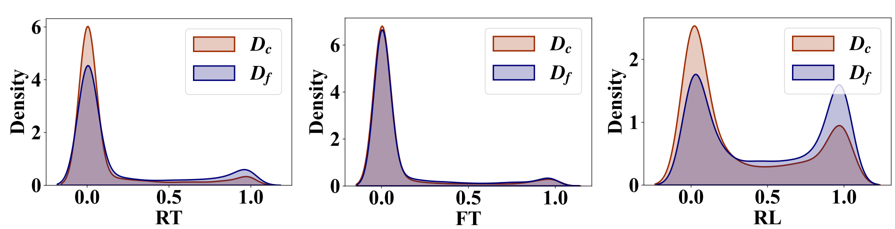


Figure 5: Non-conformity score density of calibration data $\mathcal{D}_c$ and forget data $\mathcal{D}_f$ **with our unlearning framework** in CIFAR-10 with ResNet-18 under 10% random data forgetting scenario. Our unlearning framework shifts the distribution of the forget data to the right, demonstrating improved forgetting quality.

Table 6: Unlearning performance of 9 unlearning methods on **CIFAR-10** with **ResNet-18** in 10% **random data forgetting** scenario. The results are reported in the format a±b, where a is the mean and b is the standard deviation from 3 independent trials. The performance gap relative to Retrain method is represented in (•).

Methods	α	Coverage $\mathcal{D}_f \downarrow$	Coverage $\mathcal{D}_{test} \uparrow$	Set Size $\mathcal{D}_f \uparrow$	Set Size $\mathcal{D}_{test} \downarrow$	CR $\mathcal{D}_f \downarrow$	CR $\mathcal{D}_{test} \uparrow$	$\hat{q}$
RT **UA**8.6%, **RA**99.7%, **TA**91.8%	0.05	$0.941_{\pm0.002}$(0.000)	$0.944_{\pm0.005}$(0.000)	$1.089_{\pm0.002}$(0.000)	$1.074_{\pm0.011}$(0.000)	$0.864_{\pm0.004}$(0.000)	$0.879_{\pm0.004}$(0.000)	$0.883_{\pm0.007}$
	0.1	$0.881_{\pm0.000}$(0.000)	$0.895_{\pm0.010}$(0.000)	$0.934_{\pm0.004}$(0.000)	$0.947_{\pm0.008}$(0.000)	$0.943_{\pm0.011}$(0.000)	$0.945_{\pm0.001}$(0.000)	$0.192_{\pm0.001}$
	0.15	$0.820_{\pm0.002}$(0.000)	$0.839_{\pm0.008}$(0.000)	$0.841_{\pm0.009}$(0.000)	$0.867_{\pm0.009}$(0.000)	$0.975_{\pm0.001}$(0.000)	$0.968_{\pm0.003}$(0.000)	$0.015_{\pm0.011}$
	0.2	$0.780_{\pm0.007}$(0.000)	$0.808_{\pm0.004}$(0.000)	$0.789_{\pm0.002}$(0.000)	$0.824_{\pm0.009}$(0.000)	$0.988_{\pm0.006}$(0.000)	$0.981_{\pm0.007}$(0.000)	$0.003_{\pm0.002}$
FT **UA**3.8%, **RA**98.1%, **TA**91.6%	0.05	$0.994_{\pm0.001}$(0.053)	$0.951_{\pm0.004}$(0.007)	$1.008_{\pm0.003}$(0.081)	$1.026_{\pm0.008}$(0.048)	$0.986_{\pm0.003}$(0.122)	$0.927_{\pm0.004}$(0.048)	$0.721_{\pm0.045}$
	0.1	$0.968_{\pm0.001}$(0.087)	$0.899_{\pm0.005}$(0.004)	$0.969_{\pm0.001}$(0.035)	$0.924_{\pm0.008}$(0.023)	$0.998_{\pm0.001}$(0.055)	$0.972_{\pm0.003}$(0.027)	$0.079_{\pm0.013}$
	0.15	$0.915_{\pm0.003}$(0.095)	$0.848_{\pm0.002}$(0.009)	$0.916_{\pm0.003}$(0.075)	$0.860_{\pm0.001}$(0.007)	$1.000_{\pm0.000}$(0.025)	$0.986_{\pm0.002}$(0.018)	$0.008_{\pm0.000}$
	0.2	$0.861_{\pm0.010}$(0.081)	$0.806_{\pm0.008}$(0.002)	$0.861_{\pm0.010}$(0.072)	$0.811_{\pm0.009}$(0.013)	$1.000_{\pm0.000}$(0.012)	$0.993_{\pm0.001}$(0.012)	$0.002_{\pm0.000}$
RL **UA**7.6%, **RA**97.4%, **TA**90.6%	0.05	$0.970_{\pm0.006}$(0.029)	$0.949_{\pm0.005}$(0.005)	$1.242_{\pm0.151}$(0.153)	$1.197_{\pm0.098}$(0.123)	$0.788_{\pm0.089}$(0.076)	$0.796_{\pm0.061}$(0.083)	$0.877_{\pm0.057}$
	0.1	$0.913_{\pm0.010}$(0.032)	$0.897_{\pm0.007}$(0.002)	$0.975_{\pm0.028}$(0.041)	$0.980_{\pm0.025}$(0.033)	$0.936_{\pm0.022}$(0.007)	$0.916_{\pm0.019}$(0.029)	$0.572_{\pm0.059}$
	0.15	$0.825_{\pm0.006}$(0.005)	$0.843_{\pm0.009}$(0.004)	$0.854_{\pm0.010}$(0.013)	$0.888_{\pm0.017}$(0.021)	$0.966_{\pm0.006}$(0.009)	$0.949_{\pm0.009}$(0.019)	$0.329_{\pm0.021}$
	0.2	$0.755_{\pm0.021}$(0.025)	$0.798_{\pm0.005}$(0.010)	$0.774_{\pm0.020}$(0.015)	$0.832_{\pm0.009}$(0.008)	$0.976_{\pm0.002}$(0.012)	$0.959_{\pm0.005}$(0.022)	$0.234_{\pm0.028}$
GA **UA**0.6%, **RA**99.5%, **TA**94.1%	0.05	$0.994_{\pm0.003}$(0.053)	$0.945_{\pm0.008}$(0.001)	$1.002_{\pm0.010}$(0.087)	$1.009_{\pm0.010}$(0.065)	$0.994_{\pm0.016}$(0.130)	$0.936_{\pm0.011}$(0.057)	$0.621_{\pm0.015}$
	0.1	$0.990_{\pm0.005}$(0.109)	$0.905_{\pm0.019}$(0.010)	$0.990_{\pm0.014}$(0.056)	$0.928_{\pm0.005}$(0.019)	$0.998_{\pm0.002}$(0.055)	$0.973_{\pm0.012}$(0.028)	$0.062_{\pm0.016}$
	0.15	$0.969_{\pm0.012}$(0.149)	$0.848_{\pm0.004}$(0.009)	$0.969_{\pm0.014}$(0.128)	$0.858_{\pm0.019}$(0.009)	$1.000_{\pm0.014}$(0.025)	$0.986_{\pm0.008}$(0.018)	$0.006_{\pm0.009}$
	0.2	$0.925_{\pm0.012}$(0.145)	$0.805_{\pm0.022}$(0.003)	$0.924_{\pm0.007}$(0.135)	$0.811_{\pm0.013}$(0.013)	$0.998_{\pm0.013}$(0.010)	$0.992_{\pm0.012}$(0.011)	$0.003_{\pm0.005}$
Teacher **UA**0.8%, **RA**99.4%, **TA**93.5%	0.05	$0.991_{\pm0.022}$(0.050)	$0.941_{\pm0.001}$(0.003)	$1.003_{\pm0.012}$(0.086)	$1.021_{\pm0.009}$(0.053)	$0.993_{\pm0.021}$(0.129)	$0.922_{\pm0.015}$(0.043)	$0.744_{\pm0.015}$
	0.1	$0.967_{\pm0.000}$(0.086)	$0.898_{\pm0.007}$(0.003)	$0.963_{\pm0.007}$(0.029)	$0.929_{\pm0.018}$(0.018)	$0.998_{\pm0.000}$(0.055)	$0.969_{\pm0.013}$(0.024)	$0.591_{\pm0.005}$
	0.15	$0.913_{\pm0.006}$(0.093)	$0.845_{\pm0.007}$(0.006)	$0.912_{\pm0.014}$(0.071)	$0.859_{\pm0.005}$(0.008)	$0.996_{\pm0.018}$(0.021)	$0.983_{\pm0.015}$(0.015)	$0.481_{\pm0.009}$
	0.2	$0.865_{\pm0.009}$(0.085)	$0.806_{\pm0.021}$(0.002)	$0.866_{\pm0.009}$(0.077)	$0.816_{\pm0.012}$(0.008)	$0.998_{\pm0.008}$(0.010)	$0.988_{\pm0.016}$(0.007)	$0.426_{\pm0.007}$
SSD **UA**0.5%, **RA**99.5%, **TA**94.2%	0.05	$0.996_{\pm0.004}$(0.055)	$0.945_{\pm0.002}$(0.001)	$0.999_{\pm0.019}$(0.090)	$1.008_{\pm0.011}$(0.066)	$0.994_{\pm0.006}$(0.130)	$0.936_{\pm0.014}$(0.057)	$0.622_{\pm0.019}$
	0.1	$0.987_{\pm0.003}$(0.106)	$0.902_{\pm0.010}$(0.007)	$0.990_{\pm0.003}$(0.056)	$0.926_{\pm0.017}$(0.021)	$0.998_{\pm0.020}$(0.055)	$0.973_{\pm0.002}$(0.028)	$0.063_{\pm0.022}$
	0.15	$0.967_{\pm0.016}$(0.147)	$0.849_{\pm0.009}$(0.010)	$0.965_{\pm0.000}$(0.124)	$0.862_{\pm0.012}$(0.005)	$1.002_{\pm0.019}$(0.027)	$0.990_{\pm0.002}$(0.022)	$0.007_{\pm0.007}$
	0.2	$0.922_{\pm0.006}$(0.142)	$0.803_{\pm0.000}$(0.005)	$0.923_{\pm0.009}$(0.134)	$0.811_{\pm0.005}$(0.013)	$1.002_{\pm0.020}$(0.014)	$0.992_{\pm0.009}$(0.011)	$0.001_{\pm0.005}$
NegGrad+ **UA**8.7%, **RA**98.8%, **TA**92.2%	0.05	$0.934_{\pm0.007}$(0.007)	$0.948_{\pm0.007}$(0.004)	$1.068_{\pm0.017}$(0.021)	$1.086_{\pm0.022}$(0.012)	$0.875_{\pm0.008}$(0.011)	$0.873_{\pm0.011}$(0.006)	$0.989_{\pm0.013}$
	0.1	$0.895_{\pm0.004}$(0.014)	$0.898_{\pm0.008}$(0.003)	$0.964_{\pm0.008}$(0.030)	$0.950_{\pm0.013}$(0.003)	$0.928_{\pm0.005}$(0.015)	$0.946_{\pm0.005}$(0.001)	$0.044_{\pm0.041}$
	0.15	$0.851_{\pm0.013}$(0.031)	$0.851_{\pm0.016}$(0.012)	$0.896_{\pm0.016}$(0.055)	$0.876_{\pm0.019}$(0.009)	$0.950_{\pm0.003}$(0.025)	$0.971_{\pm0.003}$(0.003)	$0.000_{\pm0.000}$
	0.2	$0.800_{\pm0.006}$(0.020)	$0.799_{\pm0.001}$(0.009)	$0.832_{\pm0.006}$(0.043)	$0.813_{\pm0.001}$(0.011)	$0.961_{\pm0.002}$(0.027)	$0.983_{\pm0.001}$(0.002)	$0.000_{\pm0.000}$
Salun **UA**3.7%, **RA**98.9%, **TA**91.8%	0.05	$0.987_{\pm0.002}$(0.046)	$0.950_{\pm0.001}$(0.006)	$1.132_{\pm0.007}$(0.043)	$1.143_{\pm0.002}$(0.069)	$0.872_{\pm0.006}$(0.008)	$0.832_{\pm0.003}$(0.047)	$0.867_{\pm0.001}$
	0.1	$0.936_{\pm0.010}$(0.055)	$0.896_{\pm0.008}$(0.001)	$0.956_{\pm0.012}$(0.022)	$0.954_{\pm0.011}$(0.007)	$0.979_{\pm0.003}$(0.036)	$0.939_{\pm0.003}$(0.006)	$0.489_{\pm0.029}$
	0.15	$0.871_{\pm0.005}$(0.051)	$0.849_{\pm0.008}$(0.010)	$0.881_{\pm0.006}$(0.040)	$0.886_{\pm0.010}$(0.019)	$0.989_{\pm0.002}$(0.014)	$0.958_{\pm0.002}$(0.010)	$0.314_{\pm0.020}$
	0.2	$0.788_{\pm0.010}$(0.008)	$0.794_{\pm0.001}$(0.014)	$0.794_{\pm0.010}$(0.005)	$0.821_{\pm0.004}$(0.003)	$0.992_{\pm0.001}$(0.004)	$0.966_{\pm0.003}$(0.015)	$0.221_{\pm0.005}$
SFRon **UA**4.8%, **RA**97.4%, **TA**91.4%	0.05	$0.977_{\pm0.003}$(0.036)	$0.953_{\pm0.004}$(0.009)	$1.100_{\pm0.023}$(0.011)	$1.143_{\pm0.021}$(0.069)	$0.889_{\pm0.015}$(0.025)	$0.834_{\pm0.012}$(0.045)	$0.926_{\pm0.018}$
	0.1	$0.945_{\pm0.004}$(0.064)	$0.905_{\pm0.005}$(0.010)	$0.986_{\pm0.005}$(0.052)	$0.977_{\pm0.008}$(0.030)	$0.958_{\pm0.001}$(0.015)	$0.927_{\pm0.003}$(0.018)	$0.435_{\pm0.043}$
	0.15	$0.895_{\pm0.002}$(0.075)	$0.847_{\pm0.002}$(0.008)	$0.912_{\pm0.004}$(0.071)	$0.879_{\pm0.001}$(0.012)	$0.982_{\pm0.002}$(0.007)	$0.963_{\pm0.003}$(0.005)	$0.082_{\pm0.007}$
	0.2	$0.857_{\pm0.008}$(0.077)	$0.808_{\pm0.002}$(0.000)	$0.868_{\pm0.007}$(0.079)	$0.826_{\pm0.005}$(0.002)	$0.988_{\pm0.002}$(0.000)	$0.978_{\pm0.004}$(0.003)	$0.025_{\pm0.005}$

Table 7: Unlearning performance of 9 unlearning methods on **CIFAR-10** with **ResNet18** in 50% **random data forgetting** scenario.

Methods	α	Coverage $\mathcal{D}_f \downarrow$	Coverage $\mathcal{D}_{test} \uparrow$	Set Size $\mathcal{D}_f \uparrow$	Set Size $\mathcal{D}_{test} \downarrow$	CR $\mathcal{D}_f \downarrow$	CR $\mathcal{D}_{test} \uparrow$	$\hat{q}$
RT **UA**11.0%, **RA**99.8%, **TA**89.2%	0.05	$0.955_{\pm 0.004}$(0.000)	$0.947_{\pm 0.005}$(0.000)	$1.287_{\pm 0.001}$(0.000)	$1.214_{\pm 0.010}$(0.000)	$0.742_{\pm 0.005}$(0.000)	$0.780_{\pm 0.006}$(0.000)	$0.984_{\pm 0.002}$
	0.1	$0.898_{\pm 0.011}$(0.000)	$0.904_{\pm 0.010}$(0.000)	$1.023_{\pm 0.005}$(0.000)	$1.021_{\pm 0.003}$(0.000)	$0.878_{\pm 0.003}$(0.000)	$0.886_{\pm 0.003}$(0.000)	$0.650_{\pm 0.004}$
	0.15	$0.833_{\pm 0.007}$(0.000)	$0.847_{\pm 0.005}$(0.000)	$0.883_{\pm 0.002}$(0.000)	$0.906_{\pm 0.003}$(0.000)	$0.943_{\pm 0.010}$(0.000)	$0.934_{\pm 0.005}$(0.000)	$0.090_{\pm 0.004}$
	0.2	$0.782_{\pm 0.005}$(0.000)	$0.814_{\pm 0.004}$(0.000)	$0.812_{\pm 0.010}$(0.000)	$0.850_{\pm 0.009}$(0.000)	$0.964_{\pm 0.005}$(0.000)	$0.958_{\pm 0.003}$(0.000)	$0.018_{\pm 0.006}$
FT **UA**2.6%, **RA**99.1%, **TA**91.8%	0.05	$0.996_{\pm 0.000}$(0.041)	$0.952_{\pm 0.002}$(0.005)	$1.007_{\pm 0.000}$(0.280)	$1.029_{\pm 0.004}$(0.185)	$0.989_{\pm 0.001}$(0.247)	$0.925_{\pm 0.002}$(0.145)	$0.738_{\pm 0.014}$
	0.1	$0.975_{\pm 0.006}$(0.077)	$0.896_{\pm 0.013}$(0.008)	$0.976_{\pm 0.006}$(0.047)	$0.921_{\pm 0.017}$(0.100)	$0.999_{\pm 0.000}$(0.121)	$0.972_{\pm 0.004}$(0.086)	$0.081_{\pm 0.033}$
	0.15	$0.936_{\pm 0.004}$(0.103)	$0.854_{\pm 0.004}$(0.007)	$0.936_{\pm 0.004}$(0.053)	$0.867_{\pm 0.006}$(0.039)	$1.000_{\pm 0.000}$(0.057)	$0.985_{\pm 0.002}$(0.051)	$0.011_{\pm 0.002}$
	0.2	$0.859_{\pm 0.010}$(0.077)	$0.790_{\pm 0.010}$(0.024)	$0.859_{\pm 0.010}$(0.047)	$0.795_{\pm 0.011}$(0.055)	$1.000_{\pm 0.000}$(0.036)	$0.993_{\pm 0.001}$(0.035)	$0.001_{\pm 0.000}$
RL **UA**10.5%, **RA**93.9%, **TA**85.8%	0.05	$0.976_{\pm 0.001}$(0.022)	$0.949_{\pm 0.002}$(0.002)	$1.973_{\pm 0.396}$(0.686)	$1.971_{\pm 0.406}$(0.757)	$0.508_{\pm 0.100}$(0.234)	$0.495_{\pm 0.098}$(0.285)	$0.899_{\pm 0.012}$
	0.1	$0.942_{\pm 0.011}$(0.043)	$0.907_{\pm 0.009}$(0.003)	$1.227_{\pm 0.103}$(0.204)	$1.235_{\pm 0.107}$(0.214)	$0.771_{\pm 0.064}$(0.107)	$0.738_{\pm 0.064}$(0.147)	$0.837_{\pm 0.016}$
	0.15	$0.891_{\pm 0.013}$(0.058)	$0.856_{\pm 0.012}$(0.009)	$1.009_{\pm 0.047}$(0.125)	$1.011_{\pm 0.045}$(0.105)	$0.884_{\pm 0.039}$(0.059)	$0.847_{\pm 0.037}$(0.087)	$0.770_{\pm 0.022}$
	0.2	$0.834_{\pm 0.003}$(0.051)	$0.799_{\pm 0.005}$(0.016)	$0.897_{\pm 0.026}$(0.086)	$0.893_{\pm 0.025}$(0.043)	$0.929_{\pm 0.024}$(0.034)	$0.895_{\pm 0.022}$(0.063)	$0.713_{\pm 0.028}$
GA **UA**0.6%, **RA**99.5%, **TA**94.3%	0.05	$0.996_{\pm 0.000}$(0.041)	$0.945_{\pm 0.008}$(0.002)	$1.003_{\pm 0.007}$(0.284)	$1.005_{\pm 0.007}$(0.209)	$1.050_{\pm 0.007}$(0.308)	$0.945_{\pm 0.007}$(0.165)	$0.616_{\pm 0.008}$
	0.1	$0.985_{\pm 0.006}$(0.087)	$0.902_{\pm 0.009}$(0.002)	$0.989_{\pm 0.006}$(0.034)	$0.926_{\pm 0.006}$(0.095)	$1.095_{\pm 0.004}$(0.217)	$0.916_{\pm 0.006}$(0.030)	$0.057_{\pm 0.005}$
	0.15	$0.966_{\pm 0.006}$(0.133)	$0.848_{\pm 0.007}$(0.001)	$0.966_{\pm 0.002}$(0.083)	$0.857_{\pm 0.009}$(0.049)	$1.141_{\pm 0.001}$(0.198)	$0.879_{\pm 0.006}$(0.055)	$0.005_{\pm 0.007}$
	0.2	$0.929_{\pm 0.004}$(0.147)	$0.809_{\pm 0.007}$(0.005)	$0.932_{\pm 0.000}$(0.120)	$0.817_{\pm 0.005}$(0.033)	$1.150_{\pm 0.002}$(0.186)	$0.871_{\pm 0.001}$(0.087)	$0.001_{\pm 0.007}$
Teacher **UA**1.6%, **RA**98.3%, **TA**91.7%	0.05	$0.985_{\pm 0.015}$(0.030)	$0.944_{\pm 0.018}$(0.003)	$1.066_{\pm 0.003}$(0.221)	$1.143_{\pm 0.012}$(0.071)	$0.923_{\pm 0.010}$(0.181)	$0.823_{\pm 0.017}$(0.043)	$0.857_{\pm 0.013}$
	0.1	$0.949_{\pm 0.012}$(0.051)	$0.909_{\pm 0.016}$(0.005)	$0.970_{\pm 0.006}$(0.053)	$0.986_{\pm 0.014}$(0.035)	$0.980_{\pm 0.001}$(0.102)	$0.918_{\pm 0.009}$(0.032)	$0.834_{\pm 0.005}$
	0.15	$0.885_{\pm 0.010}$(0.052)	$0.849_{\pm 0.018}$(0.002)	$0.894_{\pm 0.017}$(0.011)	$0.893_{\pm 0.010}$(0.013)	$0.992_{\pm 0.002}$(0.049)	$0.950_{\pm 0.013}$(0.016)	$0.813_{\pm 0.013}$
	0.2	$0.818_{\pm 0.014}$(0.036)	$0.798_{\pm 0.014}$(0.016)	$0.823_{\pm 0.009}$(0.011)	$0.826_{\pm 0.002}$(0.024)	$0.997_{\pm 0.015}$(0.033)	$0.971_{\pm 0.007}$(0.013)	$0.793_{\pm 0.012}$
SSD **UA**0.5%, **RA**99.5%, **TA**94.3%	0.05	$0.993_{\pm 0.005}$(0.038)	$0.944_{\pm 0.011}$(0.003)	$0.999_{\pm 0.007}$(0.288)	$1.001_{\pm 0.009}$(0.213)	$0.995_{\pm 0.009}$(0.253)	$0.941_{\pm 0.013}$(0.161)	$0.585_{\pm 0.014}$
	0.1	$0.991_{\pm 0.015}$(0.093)	$0.904_{\pm 0.014}$(0.000)	$0.991_{\pm 0.001}$(0.032)	$0.929_{\pm 0.011}$(0.092)	$1.000_{\pm 0.011}$(0.122)	$0.975_{\pm 0.010}$(0.089)	$0.060_{\pm 0.011}$
	0.15	$0.964_{\pm 0.016}$(0.131)	$0.850_{\pm 0.011}$(0.003)	$0.967_{\pm 0.009}$(0.084)	$0.860_{\pm 0.014}$(0.046)	$1.000_{\pm 0.001}$(0.057)	$0.988_{\pm 0.003}$(0.054)	$0.005_{\pm 0.010}$
	0.2	$0.930_{\pm 0.018}$(0.148)	$0.807_{\pm 0.002}$(0.007)	$0.929_{\pm 0.002}$(0.117)	$0.814_{\pm 0.017}$(0.036)	$1.000_{\pm 0.003}$(0.036)	$0.992_{\pm 0.001}$(0.034)	$0.002_{\pm 0.005}$
NegGrad+ **UA**2.8%, **RA**99.6%, **TA**92.9%	0.05	$0.986_{\pm 0.000}$(0.031)	$0.949_{\pm 0.001}$(0.001)	$1.039_{\pm 0.008}$(0.248)	$1.062_{\pm 0.011}$(0.152)	$0.949_{\pm 0.008}$(0.207)	$0.893_{\pm 0.008}$(0.113)	$0.855_{\pm 0.028}$
	0.1	$0.951_{\pm 0.005}$(0.053)	$0.903_{\pm 0.004}$(0.001)	$0.964_{\pm 0.008}$(0.059)	$0.944_{\pm 0.010}$(0.076)	$0.987_{\pm 0.003}$(0.109)	$0.956_{\pm 0.007}$(0.070)	$0.177_{\pm 0.055}$
	0.15	$0.889_{\pm 0.004}$(0.056)	$0.845_{\pm 0.003}$(0.002)	$0.892_{\pm 0.004}$(0.009)	$0.861_{\pm 0.003}$(0.045)	$0.996_{\pm 0.000}$(0.053)	$0.981_{\pm 0.001}$(0.047)	$0.012_{\pm 0.002}$
	0.2	$0.825_{\pm 0.003}$(0.043)	$0.796_{\pm 0.004}$(0.018)	$0.827_{\pm 0.003}$(0.015)	$0.805_{\pm 0.004}$(0.045)	$0.999_{\pm 0.000}$(0.035)	$0.989_{\pm 0.000}$(0.032)	$0.002_{\pm 0.000}$
Salun **UA**4.3%, **RA**97.7%, **TA**89.4%	0.05	$0.988_{\pm 0.001}$(0.034)	$0.951_{\pm 0.003}$(0.004)	$1.314_{\pm 0.113}$(0.027)	$1.381_{\pm 0.121}$(0.167)	$0.756_{\pm 0.064}$(0.014)	$0.692_{\pm 0.058}$(0.088)	$0.871_{\pm 0.013}$
	0.1	$0.956_{\pm 0.003}$(0.058)	$0.897_{\pm 0.005}$(0.007)	$1.015_{\pm 0.003}$(0.008)	$1.021_{\pm 0.001}$(0.001)	$0.941_{\pm 0.006}$(0.064)	$0.878_{\pm 0.004}$(0.007)	$0.776_{\pm 0.002}$
	0.15	$0.910_{\pm 0.005}$(0.078)	$0.847_{\pm 0.006}$(0.000)	$0.937_{\pm 0.009}$(0.054)	$0.916_{\pm 0.008}$(0.010)	$0.972_{\pm 0.004}$(0.029)	$0.924_{\pm 0.003}$(0.010)	$0.714_{\pm 0.010}$
	0.2	$0.856_{\pm 0.008}$(0.074)	$0.796_{\pm 0.010}$(0.019)	$0.872_{\pm 0.008}$(0.060)	$0.844_{\pm 0.008}$(0.006)	$0.982_{\pm 0.003}$(0.019)	$0.943_{\pm 0.004}$(0.015)	$0.669_{\pm 0.008}$
SFRon **UA**4.0%, **RA**97.3%, **TA**91.6%	0.05	$0.977_{\pm 0.003}$(0.022)	$0.953_{\pm 0.004}$(0.006)	$1.100_{\pm 0.023}$(0.188)	$1.143_{\pm 0.021}$(0.071)	$0.889_{\pm 0.015}$(0.147)	$0.834_{\pm 0.012}$(0.054)	$0.926_{\pm 0.018}$
	0.1	$0.945_{\pm 0.004}$(0.047)	$0.905_{\pm 0.005}$(0.001)	$0.986_{\pm 0.005}$(0.037)	$0.977_{\pm 0.008}$(0.044)	$0.958_{\pm 0.001}$(0.081)	$0.927_{\pm 0.003}$(0.042)	$0.435_{\pm 0.043}$
	0.15	$0.895_{\pm 0.002}$(0.062)	$0.847_{\pm 0.002}$(0.000)	$0.912_{\pm 0.004}$(0.029)	$0.879_{\pm 0.001}$(0.027)	$0.982_{\pm 0.002}$(0.039)	$0.963_{\pm 0.003}$(0.029)	$0.082_{\pm 0.007}$
	0.2	$0.857_{\pm 0.008}$(0.075)	$0.808_{\pm 0.002}$(0.006)	$0.868_{\pm 0.007}$(0.056)	$0.826_{\pm 0.005}$(0.024)	$0.988_{\pm 0.002}$(0.024)	$0.978_{\pm 0.004}$(0.020)	$0.025_{\pm 0.005}$

Table 8: Unlearning performance of 9 unlearning methods on **CIFAR-10** with **ResNet18** in **class-wise forgetting** scenario.

Methods	α	Coverage $\mathcal{D}_f \downarrow$	Coverage $\mathcal{D}_{tf} \downarrow$	Coverage $\mathcal{D}_{tr} \uparrow$	Set Size $\mathcal{D}_f \uparrow$	Set Size $\mathcal{D}_{tf} \uparrow$	Set Size $\mathcal{D}_{tr} \downarrow$	CR $\mathcal{D}_f \downarrow$	CR $\mathcal{D}_{tf} \downarrow$	CR $\mathcal{D}_{tr} \uparrow$	$\hat{q}_f$	$\hat{q}_{test}$
RT **UA**100%, **UA**$_{tf}$100%, **RA**99.9%, **TA**92.4%	0.05	$1.000_{\pm 0.001}$(0.000)	$1.000_{\pm 0.001}$(0.000)	$0.964_{\pm 0.008}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$1.148_{\pm 0.013}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.840_{\pm 0.002}$(0.000)	$1.000_{\pm 0.000}$	$0.982_{\pm 0.003}$
	0.1	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.001}$(0.000)	$0.882_{\pm 0.011}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$0.922_{\pm 0.009}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.100_{\pm 0.001}$(0.000)	$0.956_{\pm 0.007}$(0.000)	$1.000_{\pm 0.001}$	$0.080_{\pm 0.003}$
	0.15	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.000}$(0.000)	$0.856_{\pm 0.012}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$0.882_{\pm 0.007}$(0.000)	$0.100_{\pm 0.001}$(0.000)	$0.100_{\pm 0.001}$(0.000)	$0.970_{\pm 0.004}$(0.000)	$1.000_{\pm 0.000}$	$0.018_{\pm 0.010}$
	0.2	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.000}$(0.000)	$0.814_{\pm 0.010}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$0.830_{\pm 0.001}$(0.000)	$0.100_{\pm 0.001}$(0.000)	$0.100_{\pm 0.001}$(0.000)	$0.981_{\pm 0.002}$(0.000)	$1.000_{\pm 0.000}$	$0.003_{\pm 0.001}$
FT **UA**100%, **UA**$_{tf}$100%, **RA**96.7%, **TA**90.8%	0.05	$0.994_{\pm 0.003}$(0.006)	$0.962_{\pm 0.022}$(0.038)	$0.944_{\pm 0.011}$(0.020)	$9.854_{\pm 0.127}$(0.146)	$9.403_{\pm 0.501}$(0.597)	$1.045_{\pm 0.040}$(0.103)	$0.101_{\pm 0.001}$(0.001)	$0.102_{\pm 0.003}$(0.002)	$0.904_{\pm 0.028}$(0.065)	$1.000_{\pm 0.000}$	$0.731_{\pm 0.166}$
	0.1	$0.969_{\pm 0.011}$(0.031)	$0.882_{\pm 0.020}$(0.118)	$0.908_{\pm 0.010}$(0.026)	$9.495_{\pm 0.255}$(0.505)	$8.528_{\pm 0.571}$(1.472)	$0.956_{\pm 0.006}$(0.034)	$0.102_{\pm 0.002}$(0.002)	$0.104_{\pm 0.005}$(0.004)	$0.950_{\pm 0.007}$(0.006)	$1.000_{\pm 0.000}$	$0.314_{\pm 0.010}$
	0.15	$0.951_{\pm 0.014}$(0.049)	$0.840_{\pm 0.011}$(0.160)	$0.851_{\pm 0.031}$(0.005)	$9.265_{\pm 0.279}$(0.735)	$8.131_{\pm 0.523}$(1.869)	$0.872_{\pm 0.039}$(0.010)	$0.103_{\pm 0.003}$(0.003)	$0.103_{\pm 0.007}$(0.003)	$0.976_{\pm 0.009}$(0.006)	$1.000_{\pm 0.000}$	$0.073_{\pm 0.054}$
	0.2	$0.942_{\pm 0.014}$(0.058)	$0.818_{\pm 0.072}$(0.182)	$0.838_{\pm 0.016}$(0.023)	$9.163_{\pm 0.245}$(0.837)	$7.934_{\pm 0.533}$(2.066)	$0.854_{\pm 0.019}$(0.024)	$0.103_{\pm 0.003}$(0.003)	$0.103_{\pm 0.010}$(0.003)	$0.981_{\pm 0.005}$(0.000)	$1.000_{\pm 0.000}$	$0.039_{\pm 0.017}$
RL **UA**100%, **UA**$_{tf}$100%, **RA**98.0%, **TA**92.7%	0.05	$0.995_{\pm 0.002}$(0.005)	$0.954_{\pm 0.009}$(0.046)	$0.959_{\pm 0.015}$(0.005)	$9.993_{\pm 0.003}$(0.007)	$9.900_{\pm 0.011}$(0.100)	$1.170_{\pm 0.155}$(0.022)	$0.100_{\pm 0.000}$(0.000)	$0.096_{\pm 0.001}$(0.004)	$0.828_{\pm 0.097}$(0.012)	$1.000_{\pm 0.000}$	$0.870_{\pm 0.145}$
	0.1	$0.984_{\pm 0.003}$(0.016)	$0.907_{\pm 0.015}$(0.093)	$0.918_{\pm 0.021}$(0.036)	$9.978_{\pm 0.004}$(0.022)	$9.800_{\pm 0.019}$(0.200)	$0.982_{\pm 0.036}$(0.059)	$0.099_{\pm 0.000}$(0.001)	$0.093_{\pm 0.002}$(0.007)	$0.936_{\pm 0.022}$(0.021)	$1.000_{\pm 0.000}$	$0.469_{\pm 0.250}$
	0.15	$0.961_{\pm 0.009}$(0.039)	$0.859_{\pm 0.014}$(0.141)	$0.870_{\pm 0.019}$(0.014)	$9.950_{\pm 0.017}$(0.050)	$9.700_{\pm 0.066}$(0.300)	$0.904_{\pm 0.045}$(0.021)	$0.097_{\pm 0.001}$(0.003)	$0.089_{\pm 0.001}$(0.011)	$0.964_{\pm 0.027}$(0.006)	$1.000_{\pm 0.000}$	$0.144_{\pm 0.163}$
	0.2	$0.935_{\pm 0.027}$(0.065)	$0.815_{\pm 0.012}$(0.185)	$0.804_{\pm 0.016}$(0.010)	$9.919_{\pm 0.035}$(0.081)	$9.637_{\pm 0.076}$(0.363)	$0.820_{\pm 0.026}$(0.010)	$0.094_{\pm 0.002}$(0.006)	$0.085_{\pm 0.001}$(0.015)	$0.981_{\pm 0.012}$(0.000)	$0.999_{\pm 0.001}$	$0.014_{\pm 0.013}$
GA **UA**84.6%, **UA**$_{tf}$82.5%, **RA**96.4%, **TA**89.6%	0.05	$1.000_{\pm 0.003}$(0.000)	$1.000_{\pm 0.005}$(0.000)	$0.948_{\pm 0.004}$(0.016)	$10.000_{\pm 0.009}$(0.000)	$10.000_{\pm 0.003}$(0.000)	$1.204_{\pm 0.002}$(0.056)	$0.100_{\pm 0.007}$(0.000)	$0.100_{\pm 0.011}$(0.000)	$0.787_{\pm 0.011}$(0.053)	$1.000_{\pm 0.010}$	$0.988_{\pm 0.000}$
	0.1	$1.000_{\pm 0.003}$(0.000)	$1.000_{\pm 0.010}$(0.000)	$0.899_{\pm 0.008}$(0.017)	$10.000_{\pm 0.005}$(0.000)	$10.000_{\pm 0.006}$(0.000)	$1.005_{\pm 0.003}$(0.083)	$0.100_{\pm 0.012}$(0.000)	$0.100_{\pm 0.006}$(0.000)	$0.894_{\pm 0.002}$(0.062)	$1.000_{\pm 0.000}$	$0.562_{\pm 0.003}$
	0.15	$1.000_{\pm 0.006}$(0.000)	$1.000_{\pm 0.001}$(0.000)	$0.843_{\pm 0.011}$(0.013)	$10.000_{\pm 0.005}$(0.000)	$10.000_{\pm 0.006}$(0.000)	$0.893_{\pm 0.010}$(0.011)	$0.100_{\pm 0.004}$(0.000)	$0.100_{\pm 0.008}$(0.000)	$0.944_{\pm 0.007}$(0.026)	$1.000_{\pm 0.001}$	$0.051_{\pm 0.002}$
	0.2	$0.828_{\pm 0.003}$(0.172)	$0.782_{\pm 0.011}$(0.218)	$0.838_{\pm 0.010}$(0.024)	$9.550_{\pm 0.007}$(0.450)	$9.366_{\pm 0.002}$(0.634)	$0.884_{\pm 0.000}$(0.054)	$0.087_{\pm 0.008}$(0.013)	$0.084_{\pm 0.005}$(0.016)	$0.948_{\pm 0.010}$(0.033)	$1.000_{\pm 0.002}$	$0.038_{\pm 0.003}$
Teacher **UA**90.1%, **UA**$_{tf}$86.5%, **RA**99.5%, **TA**94.0%	0.05	$0.994_{\pm 0.003}$(0.006)	$0.959_{\pm 0.002}$(0.041)	$0.939_{\pm 0.003}$(0.025)	$9.877_{\pm 0.000}$(0.123)	$9.502_{\pm 0.003}$(0.498)	$1.000_{\pm 0.004}$(0.148)	$0.101_{\pm 0.004}$(0.001)	$0.101_{\pm 0.004}$(0.001)	$0.939_{\pm 0.001}$(0.099)	$0.955_{\pm 0.005}$	$0.588_{\pm 0.004}$
	0.1	$0.931_{\pm 0.000}$(0.069)	$0.904_{\pm 0.001}$(0.096)	$0.890_{\pm 0.001}$(0.008)	$9.199_{\pm 0.002}$(0.801)	$8.604_{\pm 0.004}$(1.396)	$0.914_{\pm 0.004}$(0.008)	$0.101_{\pm 0.004}$(0.001)	$0.105_{\pm 0.004}$(0.005)	$0.974_{\pm 0.003}$(0.018)	$0.926_{\pm 0.004}$	$0.116_{\pm 0.005}$
	0.15	$0.879_{\pm 0.004}$(0.121)	$0.881_{\pm 0.001}$(0.119)	$0.834_{\pm 0.001}$(0.022)	$8.730_{\pm 0.002}$(1.270)	$8.081_{\pm 0.001}$(1.919)	$0.845_{\pm 0.005}$(0.037)	$0.101_{\pm 0.004}$(0.001)	$0.109_{\pm 0.002}$(0.009)	$0.986_{\pm 0.004}$(0.016)	$0.921_{\pm 0.001}$	$0.017_{\pm 0.002}$
	0.2	$0.809_{\pm 0.004}$(0.191)	$0.841_{\pm 0.004}$(0.159)	$0.816_{\pm 0.000}$(0.002)	$8.141_{\pm 0.003}$(1.859)	$7.525_{\pm 0.003}$(2.475)	$0.824_{\pm 0.003}$(0.006)	$0.099_{\pm 0.002}$(0.001)	$0.112_{\pm 0.003}$(0.012)	$0.990_{\pm 0.002}$(0.009)	$0.916_{\pm 0.005}$	$0.010_{\pm 0.003}$
SSD **UA**1.16%, **UA**$_{tf}$7.75%, **RA**99.5%, **TA**94.3%	0.05	$0.995_{\pm 0.014}$(0.005)	$0.935_{\pm 0.013}$(0.065)	$0.940_{\pm 0.007}$(0.024)	$1.030_{\pm 0.014}$(8.970)	$1.067_{\pm 0.013}$(8.933)	$0.991_{\pm 0.011}$(0.157)	$0.966_{\pm 0.010}$(0.866)	$0.876_{\pm 0.007}$(0.776)	$0.949_{\pm 0.010}$(0.109)	$0.804_{\pm 0.015}$	$0.447_{\pm 0.007}$
	0.1	$0.984_{\pm 0.021}$(0.016)	$0.910_{\pm 0.009}$(0.090)	$0.880_{\pm 0.001}$(0.002)	$0.992_{\pm 0.011}$(9.008)	$0.982_{\pm 0.005}$(9.018)	$0.896_{\pm 0.003}$(0.026)	$0.992_{\pm 0.003}$(0.892)	$0.926_{\pm 0.017}$(0.826)	$0.981_{\pm 0.012}$(0.025)	$0.434_{\pm 0.007}$	$0.022_{\pm 0.005}$
	0.15	$0.960_{\pm 0.012}$(0.040)	$0.876_{\pm 0.011}$(0.124)	$0.847_{\pm 0.007}$(0.009)	$0.962_{\pm 0.007}$(9.038)	$0.931_{\pm 0.006}$(9.069)	$0.857_{\pm 0.013}$(0.025)	$0.998_{\pm 0.016}$(0.898)	$0.941_{\pm 0.002}$(0.841)	$0.989_{\pm 0.002}$(0.019)	$0.215_{\pm 0.007}$	$0.005_{\pm 0.017}$
	0.2	$0.895_{\pm 0.020}$(0.105)	$0.816_{\pm 0.010}$(0.184)	$0.823_{\pm 0.015}$(0.009)	$0.895_{\pm 0.014}$(9.105)	$0.850_{\pm 0.004}$(9.150)	$0.831_{\pm 0.002}$(0.001)	$0.999_{\pm 0.001}$(0.899)	$0.960_{\pm 0.014}$(0.860)	$0.991_{\pm 0.003}$(0.010)	$0.078_{\pm 0.003}$	$0.002_{\pm 0.009}$
NegGrad+ **UA**96.2%, **UA**$_{tf}$95.2%, **RA**97.6%, **TA**92.8%	0.05	$0.989_{\pm 0.016}$(0.011)	$0.961_{\pm 0.056}$(0.039)	$0.945_{\pm 0.027}$(0.019)	$9.432_{\pm 0.803}$(0.568)	$9.038_{\pm 1.360}$(0.962)	$1.053_{\pm 0.020}$(0.096)	$0.105_{\pm 0.007}$(0.005)	$0.107_{\pm 0.010}$(0.007)	$0.897_{\pm 0.008}$(0.058)	$1.000_{\pm 0.000}$	$0.835_{\pm 0.085}$
	0.1	$0.980_{\pm 0.029}$(0.020)	$0.954_{\pm 0.065}$(0.046)	$0.881_{\pm 0.028}$(0.001)	$9.250_{\pm 1.061}$(0.750)	$8.836_{\pm 1.647}$(1.164)	$0.913_{\pm 0.018}$(0.009)	$0.106_{\pm 0.009}$(0.006)	$0.109_{\pm 0.013}$(0.009)	$0.965_{\pm 0.012}$(0.009)	$1.000_{\pm 0.000}$	$0.057_{\pm 0.021}$
	0.15	$0.952_{\pm 0.068}$(0.048)	$0.908_{\pm 0.130}$(0.092)	$0.849_{\pm 0.026}$(0.007)	$8.600_{\pm 1.980}$(1.400)	$8.077_{\pm 2.719}$(1.923)	$0.868_{\pm 0.016}$(0.014)	$0.113_{\pm 0.018}$(0.013)	$0.116_{\pm 0.023}$(0.016)	$0.977_{\pm 0.012}$(0.007)	$1.000_{\pm 0.000}$	$0.012_{\pm 0.003}$
	0.2	$0.958_{\pm 0.060}$(0.042)	$0.921_{\pm 0.111}$(0.079)	$0.814_{\pm 0.007}$(0.001)	$8.673_{\pm 1.876}$(1.327)	$8.219_{\pm 2.519}$(1.781)	$0.828_{\pm 0.020}$(0.002)	$0.112_{\pm 0.017}$(0.012)	$0.115_{\pm 0.022}$(0.015)	$0.983_{\pm 0.015}$(0.001)	$1.000_{\pm 0.000}$	$0.004_{\pm 0.003}$
Salun **UA**100%, **UA**$_{tf}$100%, **RA**99.6%, **TA**94.3%	0.05	$0.996_{\pm 0.001}$(0.004)	$0.941_{\pm 0.008}$(0.059)	$0.952_{\pm 0.001}$(0.012)	$9.996_{\pm 0.002}$(0.004)	$9.892_{\pm 0.003}$(0.108)	$1.028_{\pm 0.008}$(0.121)	$0.100_{\pm 0.000}$(0.000)	$0.095_{\pm 0.001}$(0.005)	$0.926_{\pm 0.008}$(0.087)	$1.000_{\pm 0.000}$	$0.785_{\pm 0.049}$
	0.1	$0.988_{\pm 0.004}$(0.012)	$0.906_{\pm 0.011}$(0.094)	$0.901_{\pm 0.002}$(0.020)	$9.985_{\pm 0.003}$(0.015)	$9.817_{\pm 0.045}$(0.183)	$0.928_{\pm 0.006}$(0.006)	$0.099_{\pm 0.000}$(0.001)	$0.092_{\pm 0.001}$(0.008)	$0.971_{\pm 0.004}$(0.015)	$1.000_{\pm 0.000}$	$0.042_{\pm 0.011}$
	0.15	$0.960_{\pm 0.003}$(0.040)	$0.851_{\pm 0.005}$(0.149)	$0.878_{\pm 0.006}$(0.022)	$9.952_{\pm 0.000}$(0.048)	$9.677_{\pm 0.088}$(0.323)	$0.896_{\pm 0.005}$(0.013)	$0.096_{\pm 0.000}$(0.004)	$0.088_{\pm 0.000}$(0.012)	$0.980_{\pm 0.001}$(0.010)	$1.000_{\pm 0.000}$	$0.009_{\pm 0.001}$
	0.2	$0.915_{\pm 0.019}$(0.085)	$0.807_{\pm 0.038}$(0.193)	$0.820_{\pm 0.035}$(0.005)	$9.893_{\pm 0.024}$(0.107)	$9.511_{\pm 0.192}$(0.489)	$0.828_{\pm 0.039}$(0.002)	$0.092_{\pm 0.002}$(0.008)	$0.085_{\pm 0.002}$(0.015)	$0.990_{\pm 0.004}$(0.009)	$1.000_{\pm 0.000}$	$0.001_{\pm 0.001}$
SFRon **UA**100%, **UA**$_{tf}$100%, **RA**99.3%, **TA**94.4%	0.05	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.000}$(0.000)	$0.952_{\pm 0.005}$(0.013)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$1.022_{\pm 0.030}$(0.127)	$0.100_{\pm 0.000}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.932_{\pm 0.024}$(0.092)	$1.000_{\pm 0.000}$	$0.677_{\pm 0.206}$
	0.1	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.000}$(0.000)	$0.908_{\pm 0.013}$(0.026)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$0.937_{\pm 0.028}$(0.014)	$0.100_{\pm 0.000}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.970_{\pm 0.015}$(0.014)	$1.000_{\pm 0.000}$	$0.089_{\pm 0.092}$
	0.15	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.000}$(0.000)	$0.840_{\pm 0.026}$(0.016)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$0.849_{\pm 0.026}$(0.033)	$0.100_{\pm 0.000}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.989_{\pm 0.003}$(0.019)	$1.000_{\pm 0.000}$	$0.002_{\pm 0.001}$
	0.2	$1.000_{\pm 0.000}$(0.000)	$1.000_{\pm 0.000}$(0.000)	$0.807_{\pm 0.024}$(0.008)	$10.000_{\pm 0.000}$(0.000)	$10.000_{\pm 0.000}$(0.000)	$0.813_{\pm 0.025}$(0.017)	$0.100_{\pm 0.000}$(0.000)	$0.100_{\pm 0.000}$(0.000)	$0.992_{\pm 0.003}$(0.010)	$1.000_{\pm 0.000}$	$0.001_{\pm 0.001}$

Table 9: Unlearning performance of 9 unlearning methods on **Tiny ImageNet** with **ViT** in 10% **random data forgetting** scenario.

Methods	α	Coverage $\mathcal{D}_f \downarrow$	Coverage $\mathcal{D}_{test} \uparrow$	Set Size $\mathcal{D}_f \uparrow$	Set Size $\mathcal{D}_{test} \downarrow$	CR $\mathcal{D}_f \downarrow$	CR $\mathcal{D}_{test} \uparrow$	$\hat{q}$
RT **UA**14.7%, **RA**98.8%, **TA**86.0%	0.05	$0.944_{\pm0.006}$(0.000)	$0.949_{\pm0.026}$(0.000)	$1.876_{\pm0.009}$(0.000)	$1.840_{\pm0.014}$(0.000)	$0.503_{\pm0.018}$(0.000)	$0.516_{\pm0.018}$(0.000)	$0.984_{\pm0.002}$
	0.1	$0.892_{\pm0.006}$(0.000)	$0.900_{\pm0.025}$(0.000)	$1.151_{\pm0.002}$(0.000)	$1.144_{\pm0.018}$(0.000)	$0.775_{\pm0.016}$(0.000)	$0.786_{\pm0.026}$(0.000)	$0.853_{\pm0.003}$
	0.15	$0.841_{\pm0.024}$(0.000)	$0.850_{\pm0.017}$(0.000)	$0.956_{\pm0.014}$(0.000)	$0.956_{\pm0.017}$(0.000)	$0.880_{\pm0.014}$(0.000)	$0.889_{\pm0.019}$(0.000)	$0.539_{\pm0.001}$
	0.2	$0.790_{\pm0.015}$(0.000)	$0.799_{\pm0.023}$(0.000)	$0.846_{\pm0.004}$(0.000)	$0.854_{\pm0.014}$(0.000)	$0.934_{\pm0.012}$(0.000)	$0.935_{\pm0.015}$(0.000)	$0.238_{\pm0.012}$
FT **UA**6.9%, **RA**97.9%, **TA**84.1%	0.05	$0.994_{\pm0.005}$(0.050)	$0.950_{\pm0.019}$(0.001)	$2.133_{\pm0.008}$(0.257)	$2.440_{\pm0.011}$(0.600)	$0.466_{\pm0.009}$(0.037)	$0.389_{\pm0.016}$(0.127)	$0.994_{\pm0.020}$
	0.1	$0.978_{\pm0.007}$(0.086)	$0.903_{\pm0.003}$(0.003)	$1.234_{\pm0.010}$(0.083)	$1.317_{\pm0.001}$(0.173)	$0.792_{\pm0.018}$(0.017)	$0.685_{\pm0.001}$(0.101)	$0.935_{\pm0.012}$
	0.15	$0.938_{\pm0.001}$(0.097)	$0.851_{\pm0.010}$(0.001)	$1.014_{\pm0.005}$(0.058)	$1.017_{\pm0.016}$(0.061)	$0.925_{\pm0.007}$(0.045)	$0.836_{\pm0.016}$(0.053)	$0.681_{\pm0.003}$
	0.2	$0.888_{\pm0.009}$(0.098)	$0.801_{\pm0.012}$(0.002)	$0.915_{\pm0.006}$(0.069)	$0.885_{\pm0.000}$(0.031)	$0.970_{\pm0.020}$(0.036)	$0.905_{\pm0.005}$(0.030)	$0.326_{\pm0.011}$
RL **UA**26.9%, **RA**96.0%, **TA**81.4%	0.05	$0.969_{\pm0.021}$(0.025)	$0.952_{\pm0.008}$(0.003)	$17.890_{\pm0.003}$(16.014)	$8.572_{\pm0.010}$(6.732)	$0.054_{\pm0.013}$(0.449)	$0.111_{\pm0.002}$(0.405)	$0.996_{\pm0.019}$
	0.1	$0.892_{\pm0.017}$(0.000)	$0.902_{\pm0.013}$(0.002)	$2.639_{\pm0.017}$(1.488)	$1.843_{\pm0.019}$(0.699)	$0.338_{\pm0.022}$(0.437)	$0.489_{\pm0.013}$(0.297)	$0.971_{\pm0.014}$
	0.15	$0.793_{\pm0.021}$(0.048)	$0.855_{\pm0.008}$(0.005)	$1.225_{\pm0.013}$(0.269)	$1.164_{\pm0.000}$(0.208)	$0.648_{\pm0.002}$(0.232)	$0.734_{\pm0.000}$(0.155)	$0.894_{\pm0.022}$
	0.2	$0.681_{\pm0.010}$(0.109)	$0.803_{\pm0.003}$(0.004)	$0.831_{\pm0.006}$(0.015)	$0.946_{\pm0.011}$(0.092)	$0.820_{\pm0.022}$(0.114)	$0.849_{\pm0.006}$(0.086)	$0.715_{\pm0.013}$
GA **UA**3.2%, **RA**97.4%, **TA**84.9%	0.05	$0.996_{\pm0.003}$(0.052)	$0.947_{\pm0.002}$(0.002)	$1.539_{\pm0.004}$(0.337)	$2.018_{\pm0.007}$(0.178)	$0.647_{\pm0.003}$(0.144)	$0.469_{\pm0.002}$(0.047)	$0.988_{\pm0.004}$
	0.1	$0.986_{\pm0.006}$(0.094)	$0.900_{\pm0.000}$(0.000)	$1.104_{\pm0.006}$(0.047)	$1.224_{\pm0.005}$(0.080)	$0.894_{\pm0.003}$(0.119)	$0.736_{\pm0.006}$(0.050)	$0.899_{\pm0.001}$
	0.15	$0.967_{\pm0.002}$(0.126)	$0.852_{\pm0.005}$(0.002)	$1.003_{\pm0.008}$(0.047)	$0.993_{\pm0.004}$(0.037)	$0.964_{\pm0.005}$(0.084)	$0.859_{\pm0.006}$(0.030)	$0.632_{\pm0.009}$
	0.2	$0.934_{\pm0.001}$(0.144)	$0.800_{\pm0.007}$(0.001)	$0.946_{\pm0.008}$(0.100)	$0.871_{\pm0.008}$(0.017)	$0.987_{\pm0.008}$(0.053)	$0.919_{\pm0.005}$(0.016)	$0.296_{\pm0.009}$
Teacher **UA**17.3%, **RA**86.7%, **TA**79.0%	0.05	$0.977_{\pm0.004}$(0.033)	$0.956_{\pm0.003}$(0.007)	$5.473_{\pm0.006}$(3.597)	$5.080_{\pm0.004}$(3.240)	$0.179_{\pm0.008}$(0.324)	$0.188_{\pm0.002}$(0.328)	$0.987_{\pm0.008}$
	0.1	$0.930_{\pm0.003}$(0.038)	$0.902_{\pm0.008}$(0.002)	$1.991_{\pm0.004}$(0.840)	$1.959_{\pm0.002}$(0.815)	$0.467_{\pm0.004}$(0.308)	$0.460_{\pm0.002}$(0.326)	$0.971_{\pm0.007}$
	0.15	$0.873_{\pm0.003}$(0.032)	$0.850_{\pm0.009}$(0.000)	$1.295_{\pm0.006}$(0.339)	$1.319_{\pm0.005}$(0.363)	$0.674_{\pm0.007}$(0.206)	$0.645_{\pm0.003}$(0.244)	$0.944_{\pm0.006}$
	0.2	$0.816_{\pm0.007}$(0.026)	$0.803_{\pm0.009}$(0.004)	$1.020_{\pm0.006}$(0.174)	$1.058_{\pm0.004}$(0.204)	$0.800_{\pm0.005}$(0.134)	$0.758_{\pm0.005}$(0.177)	$0.910_{\pm0.006}$
SSD **UA**1.5%, **RA**98.5%, **TA**86.1%	0.05	$0.998_{\pm0.004}$(0.054)	$0.950_{\pm0.006}$(0.001)	$1.354_{\pm0.008}$(0.522)	$1.827_{\pm0.002}$(0.013)	$0.737_{\pm0.008}$(0.234)	$0.520_{\pm0.008}$(0.004)	$0.985_{\pm0.005}$
	0.1	$0.993_{\pm0.008}$(0.101)	$0.897_{\pm0.008}$(0.003)	$1.039_{\pm0.002}$(0.112)	$1.134_{\pm0.008}$(0.010)	$0.956_{\pm0.007}$(0.181)	$0.791_{\pm0.002}$(0.005)	$0.852_{\pm0.001}$
	0.15	$0.981_{\pm0.005}$(0.140)	$0.853_{\pm0.001}$(0.003)	$0.993_{\pm0.001}$(0.037)	$0.962_{\pm0.005}$(0.006)	$0.988_{\pm0.004}$(0.108)	$0.887_{\pm0.004}$(0.002)	$0.542_{\pm0.007}$
	0.2	$0.956_{\pm0.002}$(0.166)	$0.805_{\pm0.003}$(0.006)	$0.960_{\pm0.003}$(0.114)	$0.864_{\pm0.009}$(0.010)	$0.996_{\pm0.005}$(0.062)	$0.932_{\pm0.002}$(0.003)	$0.249_{\pm0.006}$
NegGrad+ **UA**19.4%, **RA**98.3%, **TA**84.0%	0.05	$0.999_{\pm0.000}$(0.055)	$0.890_{\pm0.002}$(0.059)	$0.949_{\pm0.002}$(0.927)	$1.614_{\pm0.023}$(0.227)	$2.184_{\pm0.052}$(1.681)	$2.499_{\pm0.059}$(1.984)	$0.995_{\pm0.000}$
	0.1	$0.995_{\pm0.001}$(0.103)	$0.848_{\pm0.000}$(0.052)	$0.898_{\pm0.000}$(0.253)	$1.093_{\pm0.005}$(0.051)	$1.225_{\pm0.007}$(0.450)	$1.287_{\pm0.003}$(0.501)	$0.933_{\pm0.002}$
	0.15	$0.987_{\pm0.000}$(0.146)	$0.814_{\pm0.001}$(0.036)	$0.850_{\pm0.001}$(0.106)	$1.009_{\pm0.000}$(0.053)	$1.017_{\pm0.002}$(0.137)	$1.023_{\pm0.003}$(0.133)	$0.685_{\pm0.002}$
	0.2	$0.966_{\pm0.001}$(0.176)	$0.783_{\pm0.003}$(0.016)	$0.802_{\pm0.002}$(0.044)	$0.972_{\pm0.000}$(0.118)	$0.922_{\pm0.004}$(0.012)	$0.891_{\pm0.001}$(0.043)	$0.320_{\pm0.001}$
Salun **UA**9.2%, **RA**97.7%, **TA**83.6%	0.05	$0.995_{\pm0.003}$(0.051)	$0.964_{\pm0.026}$(0.015)	$2.803_{\pm1.607}$(0.927)	$2.726_{\pm0.727}$(0.886)	$1.311_{\pm1.810}$(0.808)	$1.157_{\pm1.481}$(0.641)	$0.988_{\pm0.001}$
	0.1	$0.977_{\pm0.014}$(0.085)	$0.924_{\pm0.040}$(0.024)	$1.229_{\pm0.286}$(0.078)	$1.281_{\pm0.120}$(0.137)	$0.918_{\pm0.387}$(0.143)	$0.884_{\pm0.374}$(0.097)	$0.939_{\pm0.005}$
	0.15	$0.936_{\pm0.041}$(0.095)	$0.874_{\pm0.041}$(0.024)	$0.972_{\pm0.103}$(0.016)	$1.032_{\pm0.005}$(0.076)	$0.935_{\pm0.087}$(0.055)	$0.893_{\pm0.124}$(0.004)	$0.819_{\pm0.003}$
	0.2	$0.870_{\pm0.081}$(0.080)	$0.810_{\pm0.017}$(0.011)	$0.845_{\pm0.036}$(0.001)	$0.925_{\pm0.046}$(0.071)	$0.924_{\pm0.047}$(0.009)	$0.894_{\pm0.006}$(0.041)	$0.630_{\pm0.003}$
SFRon **UA**9.3%, **RA**97.0%, **TA**83.9%	0.05	$0.989_{\pm0.001}$(0.045)	$0.948_{\pm0.001}$(0.001)	$2.000_{\pm0.059}$(0.124)	$2.208_{\pm0.037}$(0.368)	$0.495_{\pm0.014}$(0.008)	$0.429_{\pm0.007}$(0.086)	$0.986_{\pm0.000}$
	0.1	$0.960_{\pm0.003}$(0.068)	$0.899_{\pm0.002}$(0.001)	$1.227_{\pm0.017}$(0.076)	$1.268_{\pm0.007}$(0.123)	$0.783_{\pm0.010}$(0.008)	$0.709_{\pm0.003}$(0.077)	$0.902_{\pm0.003}$
	0.15	$0.917_{\pm0.002}$(0.076)	$0.849_{\pm0.002}$(0.001)	$1.024_{\pm0.006}$(0.068)	$1.015_{\pm0.005}$(0.059)	$0.896_{\pm0.007}$(0.016)	$0.837_{\pm0.004}$(0.053)	$0.689_{\pm0.012}$
	0.2	$0.866_{\pm0.006}$(0.076)	$0.802_{\pm0.003}$(0.003)	$0.916_{\pm0.004}$(0.070)	$0.892_{\pm0.005}$(0.037)	$0.946_{\pm0.002}$(0.012)	$0.899_{\pm0.003}$(0.036)	$0.426_{\pm0.018}$

Table 10: Unlearning performance of 9 unlearning methods on **Tiny ImageNet** with **ViT** in 50% **random data forgetting** scenario.

Methods	α	Coverage		Set Size		CR		
		$\mathcal{D}_f \downarrow$	$\mathcal{D}_{test} \uparrow$	$\mathcal{D}_f \uparrow$	$\mathcal{D}_{test} \downarrow$	$\mathcal{D}_f \downarrow$	$\mathcal{D}_{test} \uparrow$	$\hat{q}$
RT	0.05	$0.946_{\pm0.001}$(0.000)	$0.948_{\pm0.003}$(0.000)	$2.146_{\pm0.006}$(0.000)	$2.106_{\pm0.002}$(0.000)	$0.441_{\pm0.004}$(0.000)	$0.450_{\pm0.005}$(0.000)	$0.987_{\pm0.004}$
UA16.0%, **RA**98.8%, **TA**84.9%	0.1	$0.892_{\pm0.007}$(0.000)	$0.899_{\pm0.008}$(0.000)	$1.222_{\pm0.002}$(0.000)	$1.211_{\pm0.007}$(0.000)	$0.730_{\pm0.004}$(0.000)	$0.742_{\pm0.002}$(0.000)	$0.889_{\pm0.009}$
	0.15	$0.838_{\pm0.004}$(0.000)	$0.847_{\pm0.001}$(0.000)	$0.977_{\pm0.002}$(0.000)	$0.977_{\pm0.006}$(0.000)	$0.858_{\pm0.008}$(0.000)	$0.868_{\pm0.006}$(0.000)	$0.607_{\pm0.001}$
	0.2	$0.786_{\pm0.005}$(0.000)	$0.796_{\pm0.002}$(0.000)	$0.856_{\pm0.007}$(0.000)	$0.863_{\pm0.001}$(0.000)	$0.918_{\pm0.007}$(0.000)	$0.922_{\pm0.008}$(0.000)	$0.304_{\pm0.008}$
FT	0.05	$0.995_{\pm0.013}$(0.051)	$0.949_{\pm0.024}$(0.000)	$1.879_{\pm0.014}$(0.003)	$2.216_{\pm0.003}$(0.376)	$0.527_{\pm0.028}$(0.024)	$0.428_{\pm0.020}$(0.088)	$0.992_{\pm0.019}$
UA5.4%, **RA**97.1%, **TA**84.4%	0.1	$0.979_{\pm0.021}$(0.087)	$0.901_{\pm0.014}$(0.001)	$1.183_{\pm0.018}$(0.032)	$1.281_{\pm0.020}$(0.137)	$0.828_{\pm0.029}$(0.053)	$0.701_{\pm0.010}$(0.085)	$0.926_{\pm0.025}$
	0.15	$0.953_{\pm0.024}$(0.112)	$0.850_{\pm0.022}$(0.000)	$1.014_{\pm0.011}$(0.058)	$1.017_{\pm0.026}$(0.061)	$0.940_{\pm0.027}$(0.060)	$0.839_{\pm0.004}$(0.050)	$0.681_{\pm0.020}$
	0.2	$0.910_{\pm0.029}$(0.120)	$0.806_{\pm0.024}$(0.007)	$0.937_{\pm0.018}$(0.091)	$0.895_{\pm0.001}$(0.041)	$0.977_{\pm0.029}$(0.043)	$0.902_{\pm0.007}$(0.033)	$0.345_{\pm0.016}$
RL	0.05	$0.974_{\pm0.011}$(0.028)	$0.953_{\pm0.001}$(0.005)	$26.032_{\pm0.007}$(23.886)	$23.369_{\pm0.008}$(21.263)	$0.038_{\pm0.015}$(0.403)	$0.038_{\pm0.016}$(0.412)	$0.994_{\pm0.010}$
UA22.5%, **RA**93.5%, **TA**77.1%	0.1	$0.930_{\pm0.016}$(0.038)	$0.902_{\pm0.013}$(0.003)	$5.277_{\pm0.001}$(4.055)	$4.621_{\pm0.007}$(3.410)	$0.178_{\pm0.011}$(0.552)	$0.197_{\pm0.001}$(0.545)	$0.987_{\pm0.008}$
	0.15	$0.875_{\pm0.011}$(0.037)	$0.856_{\pm0.008}$(0.009)	$1.758_{\pm0.004}$(0.781)	$1.657_{\pm0.005}$(0.680)	$0.496_{\pm0.006}$(0.362)	$0.516_{\pm0.009}$(0.352)	$0.970_{\pm0.017}$
	0.2	$0.810_{\pm0.006}$(0.024)	$0.805_{\pm0.013}$(0.009)	$1.147_{\pm0.005}$(0.291)	$1.144_{\pm0.005}$(0.281)	$0.707_{\pm0.004}$(0.211)	$0.707_{\pm0.013}$(0.215)	$0.945_{\pm0.005}$
GA	0.05	$0.998_{\pm0.007}$(0.052)	$0.949_{\pm0.001}$(0.001)	$1.807_{\pm0.001}$(0.339)	$2.338_{\pm0.001}$(0.232)	$0.552_{\pm0.006}$(0.111)	$0.407_{\pm0.006}$(0.043)	$0.992_{\pm0.006}$
UA3.9%, **RA**96.1%, **TA**84.2%	0.1	$0.986_{\pm0.009}$(0.094)	$0.896_{\pm0.007}$(0.003)	$1.147_{\pm0.003}$(0.075)	$1.278_{\pm0.007}$(0.067)	$0.863_{\pm0.008}$(0.133)	$0.703_{\pm0.002}$(0.039)	$0.918_{\pm0.010}$
	0.15	$0.968_{\pm0.008}$(0.130)	$0.850_{\pm0.002}$(0.003)	$1.015_{\pm0.008}$(0.038)	$1.020_{\pm0.002}$(0.043)	$0.954_{\pm0.009}$(0.096)	$0.835_{\pm0.002}$(0.033)	$0.696_{\pm0.009}$
	0.2	$0.931_{\pm0.011}$(0.145)	$0.804_{\pm0.004}$(0.008)	$0.948_{\pm0.000}$(0.092)	$0.893_{\pm0.003}$(0.030)	$0.983_{\pm0.006}$(0.065)	$0.900_{\pm0.004}$(0.022)	$0.363_{\pm0.002}$
Teacher	0.05	$0.967_{\pm0.013}$(0.021)	$0.950_{\pm0.017}$(0.002)	$6.465_{\pm0.007}$(4.319)	$6.233_{\pm0.004}$(4.127)	$0.151_{\pm0.002}$(0.290)	$0.151_{\pm0.006}$(0.299)	$0.990_{\pm0.014}$
UA22.1%, **RA**85.7%, **TA**76.2%	0.1	$0.922_{\pm0.008}$(0.030)	$0.899_{\pm0.002}$(0.000)	$2.202_{\pm0.012}$(0.980)	$2.167_{\pm0.005}$(0.956)	$0.418_{\pm0.009}$(0.312)	$0.419_{\pm0.024}$(0.323)	$0.977_{\pm0.001}$
	0.15	$0.869_{\pm0.025}$(0.031)	$0.852_{\pm0.002}$(0.005)	$1.467_{\pm0.015}$(0.490)	$1.459_{\pm0.004}$(0.482)	$0.591_{\pm0.005}$(0.267)	$0.581_{\pm0.001}$(0.287)	$0.958_{\pm0.021}$
	0.2	$0.814_{\pm0.020}$(0.028)	$0.801_{\pm0.017}$(0.005)	$1.125_{\pm0.005}$(0.269)	$1.138_{\pm0.001}$(0.275)	$0.718_{\pm0.017}$(0.200)	$0.704_{\pm0.009}$(0.218)	$0.927_{\pm0.017}$
SSD	0.05	$0.999_{\pm0.001}$(0.053)	$0.952_{\pm0.001}$(0.004)	$1.346_{\pm0.001}$(0.800)	$1.824_{\pm0.000}$(0.282)	$0.742_{\pm0.000}$(0.301)	$0.522_{\pm0.001}$(0.072)	$0.986_{\pm0.001}$
UA1.3%, **RA**98.4%, **TA**86.1%	0.1	$0.995_{\pm0.001}$(0.103)	$0.897_{\pm0.000}$(0.002)	$1.033_{\pm0.001}$(0.189)	$1.135_{\pm0.001}$(0.076)	$0.959_{\pm0.000}$(0.229)	$0.790_{\pm0.000}$(0.048)	$0.847_{\pm0.001}$
	0.15	$0.982_{\pm0.001}$(0.144)	$0.847_{\pm0.000}$(0.000)	$0.987_{\pm0.000}$(0.010)	$0.956_{\pm0.000}$(0.021)	$0.989_{\pm0.001}$(0.131)	$0.890_{\pm0.001}$(0.022)	$0.517_{\pm0.001}$
	0.2	$0.959_{\pm0.001}$(0.173)	$0.804_{\pm0.001}$(0.008)	$0.961_{\pm0.000}$(0.105)	$0.862_{\pm0.000}$(0.001)	$0.995_{\pm0.001}$(0.077)	$0.932_{\pm0.001}$(0.010)	$0.243_{\pm0.001}$
NegGrad+	0.05	$0.999_{\pm0.000}$(0.053)	$0.979_{\pm0.001}$(0.031)	$0.946_{\pm0.002}$(1.200)	$1.443_{\pm0.028}$(0.663)	$2.248_{\pm0.063}$(1.807)	$2.358_{\pm0.095}$(1.908)	$0.992_{\pm0.001}$
UA11.5%, **RA**98.7%, **TA**83.8%	0.1	$0.996_{\pm0.000}$(0.104)	$0.946_{\pm0.002}$(0.047)	$0.900_{\pm0.003}$(0.322)	$1.078_{\pm0.006}$(0.134)	$1.295_{\pm0.010}$(0.565)	$1.332_{\pm0.008}$(0.590)	$0.933_{\pm0.003}$
	0.15	$0.990_{\pm0.000}$(0.152)	$0.900_{\pm0.003}$(0.052)	$0.853_{\pm0.004}$(0.124)	$1.008_{\pm0.002}$(0.031)	$1.032_{\pm0.010}$(0.174)	$1.033_{\pm0.011}$(0.165)	$0.712_{\pm0.015}$
	0.2	$0.977_{\pm0.000}$(0.191)	$0.848_{\pm0.003}$(0.052)	$0.805_{\pm0.002}$(0.052)	$0.982_{\pm0.000}$(0.119)	$0.909_{\pm0.004}$(0.009)	$0.898_{\pm0.007}$(0.024)	$0.381_{\pm0.009}$
Salun	0.05	$0.993_{\pm0.003}$(0.047)	$0.962_{\pm0.026}$(0.014)	$3.284_{\pm2.048}$(1.138)	$4.112_{\pm0.813}$(2.007)	$1.546_{\pm2.290}$(1.105)	$1.558_{\pm2.336}$(1.108)	$0.989_{\pm0.001}$
UA9.2%, **RA**95.7%, **TA**81.9%	0.1	$0.976_{\pm0.011}$(0.084)	$0.924_{\pm0.039}$(0.026)	$1.386_{\pm0.423}$(0.164)	$1.579_{\pm0.130}$(0.368)	$0.922_{\pm0.566}$(0.192)	$0.896_{\pm0.607}$(0.154)	$0.973_{\pm0.002}$
	0.15	$0.944_{\pm0.024}$(0.106)	$0.876_{\pm0.046}$(0.029)	$1.051_{\pm0.175}$(0.074)	$1.139_{\pm0.017}$(0.162)	$0.919_{\pm0.194}$(0.061)	$0.871_{\pm0.226}$(0.003)	$0.942_{\pm0.002}$
	0.2	$0.900_{\pm0.044}$(0.114)	$0.825_{\pm0.049}$(0.029)	$0.910_{\pm0.097}$(0.054)	$0.969_{\pm0.037}$(0.105)	$0.928_{\pm0.040}$(0.011)	$0.876_{\pm0.063}$(0.045)	$0.893_{\pm0.002}$
SFRon	0.05	$0.994_{\pm0.001}$(0.048)	$0.947_{\pm0.003}$(0.001)	$2.010_{\pm0.188}$(0.136)	$2.327_{\pm0.087}$(0.222)	$0.497_{\pm0.045}$(0.057)	$0.407_{\pm0.016}$(0.043)	$0.983_{\pm0.002}$
UA6.3%, **RA**96.8%, **TA**82.9%	0.1	$0.980_{\pm0.006}$(0.087)	$0.900_{\pm0.003}$(0.001)	$1.245_{\pm0.060}$(0.023)	$1.338_{\pm0.039}$(0.126)	$0.788_{\pm0.041}$(0.058)	$0.673_{\pm0.020}$(0.069)	$0.909_{\pm0.003}$
	0.15	$0.951_{\pm0.011}$(0.113)	$0.849_{\pm0.003}$(0.001)	$1.041_{\pm0.020}$(0.065)	$1.044_{\pm0.023}$(0.067)	$0.913_{\pm0.028}$(0.055)	$0.813_{\pm0.016}$(0.055)	$0.738_{\pm0.029}$
	0.2	$0.910_{\pm0.011}$(0.125)	$0.803_{\pm0.003}$(0.008)	$0.947_{\pm0.006}$(0.091)	$0.910_{\pm0.022}$(0.046)	$0.961_{\pm0.017}$(0.044)	$0.884_{\pm0.017}$(0.038)	$0.523_{\pm0.068}$

Table 11: Unlearning performance of 9 unlearning methods on **Tiny ImageNet** with **ViT** in **class-wise forgetting** scenario.

Methods	α	Coverage			Set Size			CR				
		$\mathcal{D}_f \downarrow$	$\mathcal{D}_{tf} \downarrow$	$\mathcal{D}_{tr} \uparrow$	$\mathcal{D}_f \uparrow$	$\mathcal{D}_{tf} \uparrow$	$\mathcal{D}_{tr} \downarrow$	$\mathcal{D}_f \downarrow$	$\mathcal{D}_{tf} \downarrow$	$\mathcal{D}_{tr} \uparrow$	$\hat{q}_f$	$\hat{q}_{test}$
RT	0.05	$1.000_{\pm0.000}$(0.000)	$1.000_{\pm0.000}$(0.000)	$0.950_{\pm0.003}$(0.000)	$200.000_{\pm0.000}$(0.000)	$200.000_{\pm0.000}$(0.000)	$1.785_{\pm0.056}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.532_{\pm0.009}$(0.000)	$1.000_{\pm0.000}$	$0.984_{\pm0.002}$
UA100%, **UA**$_{tf}$100%,	0.1	$0.936_{\pm0.011}$(0.000)	$0.960_{\pm0.016}$(0.000)	$0.903_{\pm0.009}$(0.000)	$192.882_{\pm0.912}$(0.000)	$193.340_{\pm2.620}$(0.000)	$1.146_{\pm0.002}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.788_{\pm0.008}$(0.000)	$1.000_{\pm0.000}$	$0.859_{\pm0.004}$
RA98.7%, **TA**86.4%	0.15	$0.904_{\pm0.039}$(0.000)	$0.960_{\pm0.046}$(0.000)	$0.853_{\pm0.005}$(0.000)	$186.791_{\pm2.173}$(0.000)	$188.880_{\pm1.802}$(0.000)	$0.957_{\pm0.010}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.892_{\pm0.003}$(0.000)	$1.000_{\pm0.000}$	$0.535_{\pm0.002}$
	0.2	$0.787_{\pm0.061}$(0.000)	$0.860_{\pm0.024}$(0.000)	$0.805_{\pm0.003}$(0.000)	$171.051_{\pm3.183}$(0.000)	$174.480_{\pm2.311}$(0.000)	$0.860_{\pm0.010}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.936_{\pm0.002}$(0.000)	$1.000_{\pm0.000}$	$0.232_{\pm0.001}$
FT	0.05	$0.993_{\pm0.006}$(0.007)	$0.960_{\pm0.009}$(0.040)	$0.952_{\pm0.006}$(0.002)	$8.360_{\pm0.007}$(191.640)	$8.280_{\pm0.006}$(191.720)	$2.442_{\pm0.011}$(0.657)	$0.119_{\pm0.018}$(0.114)	$0.116_{\pm0.001}$(0.111)	$0.390_{\pm0.023}$(0.142)	$0.999_{\pm0.006}$	$0.993_{\pm0.005}$
UA13.8%, **UA**$_{tf}$22.0%,	0.1	$0.984_{\pm0.009}$(0.048)	$0.860_{\pm0.013}$(0.100)	$0.898_{\pm0.005}$(0.005)	$1.802_{\pm0.009}$(191.080)	$1.660_{\pm0.018}$(191.680)	$1.287_{\pm0.009}$(0.141)	$0.546_{\pm0.008}$(0.541)	$0.518_{\pm0.004}$(0.513)	$0.698_{\pm0.019}$(0.090)	$0.971_{\pm0.019}$	$0.924_{\pm0.016}$
RA97.5%, **TA**84.1%	0.15	$0.902_{\pm0.019}$(0.002)	$0.800_{\pm0.004}$(0.160)	$0.852_{\pm0.017}$(0.001)	$1.120_{\pm0.021}$(185.671)	$1.040_{\pm0.006}$(187.840)	$1.021_{\pm0.017}$(0.064)	$0.806_{\pm0.012}$(0.801)	$0.769_{\pm0.013}$(0.764)	$0.835_{\pm0.022}$(0.057)	$0.809_{\pm0.010}$	$0.686_{\pm0.004}$
	0.2	$0.860_{\pm0.021}$(0.073)	$0.760_{\pm0.003}$(0.100)	$0.800_{\pm0.018}$(0.005)	$0.969_{\pm0.002}$(170.082)	$0.960_{\pm0.003}$(173.520)	$0.882_{\pm0.010}$(0.022)	$0.888_{\pm0.005}$(0.883)	$0.792_{\pm0.002}$(0.787)	$0.907_{\pm0.006}$(0.029)	$0.595_{\pm0.002}$	$0.338_{\pm0.019}$
RL	0.05	$0.998_{\pm0.005}$(0.002)	$0.980_{\pm0.003}$(0.020)	$0.952_{\pm0.049}$(0.002)	$199.489_{\pm0.512}$(0.511)	$195.220_{\pm1.003}$(4.780)	$2.317_{\pm0.009}$(0.532)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.411_{\pm0.000}$(0.121)	$1.000_{\pm0.000}$	$0.995_{\pm0.032}$
UA100%, **UA**$_{tf}$100%,	0.1	$0.971_{\pm0.013}$(0.035)	$0.900_{\pm0.017}$(0.060)	$0.900_{\pm0.002}$(0.003)	$180.442_{\pm0.710}$(12.440)	$170.960_{\pm0.948}$(22.380)	$1.237_{\pm0.050}$(0.991)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.727_{\pm0.016}$(0.061)	$1.000_{\pm0.000}$	$0.925_{\pm0.024}$
RA98.2%, **TA**84.6%	0.15	$0.922_{\pm0.011}$(0.018)	$0.900_{\pm0.011}$(0.060)	$0.852_{\pm0.015}$(0.001)	$165.884_{\pm2.037}$(20.907)	$159.980_{\pm1.012}$(28.900)	$1.001_{\pm0.003}$(0.044)	$0.006_{\pm0.001}$(0.001)	$0.006_{\pm0.000}$(0.001)	$0.851_{\pm0.023}$(0.041)	$1.000_{\pm0.000}$	$0.641_{\pm0.035}$
	0.2	$0.882_{\pm0.007}$(0.095)	$0.860_{\pm0.007}$(0.000)	$0.807_{\pm0.007}$(0.002)	$154.896_{\pm2.028}$(16.155)	$149.280_{\pm3.013}$(25.200)	$0.886_{\pm0.032}$(0.026)	$0.006_{\pm0.000}$(0.001)	$0.006_{\pm0.001}$(0.001)	$0.912_{\pm0.013}$(0.024)	$1.000_{\pm0.000}$	$0.262_{\pm0.022}$
GA	0.05	$1.000_{\pm0.001}$(0.000)	$0.980_{\pm0.002}$(0.020)	$0.948_{\pm0.026}$(0.002)	$22.836_{\pm0.045}$(177.164)	$20.600_{\pm0.011}$(179.400)	$1.781_{\pm0.017}$(0.004)	$0.044_{\pm0.017}$(0.019)	$0.048_{\pm0.028}$(0.043)	$0.532_{\pm0.013}$(0.000)	$1.000_{\pm0.000}$	$0.984_{\pm0.033}$
UA9.1%, **UA**$_{tf}$20.0%,	0.1	$0.991_{\pm0.022}$(0.055)	$0.900_{\pm0.014}$(0.060)	$0.897_{\pm0.016}$(0.006)	$1.631_{\pm0.031}$(191.251)	$1.720_{\pm0.005}$(191.620)	$1.133_{\pm0.044}$(0.013)	$0.608_{\pm0.006}$(0.603)	$0.523_{\pm0.007}$(0.518)	$0.792_{\pm0.037}$(0.004)	$0.972_{\pm0.033}$	$0.849_{\pm0.039}$
RA98.6%, **TA**86.1%	0.15	$0.958_{\pm0.002}$(0.054)	$0.820_{\pm0.010}$(0.140)	$0.850_{\pm0.006}$(0.003)	$1.151_{\pm0.039}$(185.640)	$1.140_{\pm0.042}$(187.740)	$0.958_{\pm0.026}$(0.001)	$0.832_{\pm0.003}$(0.827)	$0.719_{\pm0.021}$(0.714)	$0.887_{\pm0.044}$(0.005)	$0.868_{\pm0.023}$	$0.535_{\pm0.011}$
	0.2	$0.880_{\pm0.047}$(0.093)	$0.800_{\pm0.051}$(0.060)	$0.803_{\pm0.025}$(0.002)	$0.929_{\pm0.002}$(170.122)	$0.900_{\pm0.009}$(173.580)	$0.861_{\pm0.006}$(0.001)	$0.947_{\pm0.036}$(0.942)	$0.889_{\pm0.029}$(0.884)	$0.933_{\pm0.027}$(0.003)	$0.473_{\pm0.016}$	$0.238_{\pm0.000}$
Teacher	0.05	$0.982_{\pm0.014}$(0.018)	$1.000_{\pm0.007}$(0.000)	$0.952_{\pm0.025}$(0.002)	$199.971_{\pm0.009}$(0.029)	$200.000_{\pm0.000}$(0.000)	$5.095_{\pm0.020}$(3.310)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.187_{\pm0.008}$(0.345)	$1.000_{\pm0.000}$	$0.989_{\pm0.001}$
UA100%, **UA**$_{tf}$100%,	0.1	$0.909_{\pm0.013}$(0.027)	$0.940_{\pm0.015}$(0.020)	$0.903_{\pm0.032}$(0.000)	$199.813_{\pm0.009}$(6.931)	$199.900_{\pm0.013}$(6.560)	$2.033_{\pm0.031}$(0.887)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.444_{\pm0.006}$(0.344)	$1.000_{\pm0.000}$	$0.965_{\pm0.003}$
RA88.8%, **TA**78.6%	0.15	$0.887_{\pm0.014}$(0.017)	$0.880_{\pm0.011}$(0.080)	$0.854_{\pm0.003}$(0.001)	$199.667_{\pm0.030}$(12.876)	$199.760_{\pm0.026}$(10.880)	$1.331_{\pm0.012}$(0.374)	$0.004_{\pm0.000}$(0.001)	$0.004_{\pm0.001}$(0.001)	$0.641_{\pm0.010}$(0.251)	$1.000_{\pm0.000}$	$0.919_{\pm0.001}$
	0.2	$0.838_{\pm0.022}$(0.051)	$0.840_{\pm0.002}$(0.020)	$0.799_{\pm0.017}$(0.006)	$199.413_{\pm0.024}$(28.362)	$199.620_{\pm0.030}$(25.140)	$1.022_{\pm0.017}$(0.162)	$0.004_{\pm0.001}$(0.001)	$0.004_{\pm0.001}$(0.001)	$0.781_{\pm0.019}$(0.155)	$1.000_{\pm0.000}$	$0.825_{\pm0.002}$
SSD	0.05	$1.000_{\pm0.000}$(0.000)	$1.000_{\pm0.000}$(0.000)	$0.950_{\pm0.017}$(0.000)	$198.769_{\pm0.052}$(1.231)	$197.320_{\pm1.010}$(2.680)	$1.866_{\pm0.019}$(0.081)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.509_{\pm0.013}$(0.023)	$1.000_{\pm0.000}$	$0.986_{\pm0.006}$
UA100%, **UA**$_{tf}$100%,	0.1	$0.949_{\pm0.017}$(0.013)	$0.900_{\pm0.012}$(0.060)	$0.897_{\pm0.007}$(0.006)	$171.073_{\pm0.209}$(21.809)	$169.360_{\pm2.002}$(23.980)	$1.141_{\pm0.014}$(0.005)	$0.006_{\pm0.000}$(0.001)	$0.005_{\pm0.000}$(0.000)	$0.786_{\pm0.021}$(0.002)	$1.000_{\pm0.000}$	$0.854_{\pm0.006}$
RA98.4%, **TA**86.1%	0.15	$0.913_{\pm0.007}$(0.009)	$0.880_{\pm0.020}$(0.080)	$0.852_{\pm0.015}$(0.001)	$157.140_{\pm1.209}$(29.651)	$154.960_{\pm0.907}$(33.920)	$0.959_{\pm0.011}$(0.002)	$0.006_{\pm0.001}$(0.001)	$0.006_{\pm0.000}$(0.001)	$0.888_{\pm0.012}$(0.004)	$1.000_{\pm0.000}$	$0.538_{\pm0.007}$
	0.2	$0.833_{\pm0.007}$(0.046)	$0.800_{\pm0.013}$(0.060)	$0.806_{\pm0.022}$(0.001)	$136.502_{\pm3.022}$(34.549)	$136.420_{\pm2.422}$(38.060)	$0.864_{\pm0.002}$(0.004)	$0.006_{\pm0.000}$(0.001)	$0.006_{\pm0.000}$(0.001)	$0.932_{\pm0.015}$(0.004)	$1.000_{\pm0.000}$	$0.254_{\pm0.005}$
NegGrad+	0.05	$1.000_{\pm0.000}$(0.000)	$1.000_{\pm0.000}$(0.000)	$0.947_{\pm0.002}$(0.003)	$200.000_{\pm0.000}$(0.000)	$200.000_{\pm0.000}$(0.000)	$1.850_{\pm0.036}$(0.065)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.512_{\pm0.009}$(0.020)	$1.000_{\pm0.000}$	$0.987_{\pm0.001}$
UA100%, **UA**$_{tf}$100%,	0.1	$0.927_{\pm0.104}$(0.009)	$0.950_{\pm0.071}$(0.010)	$0.894_{\pm0.001}$(0.009)	$193.994_{\pm8.493}$(1.112)	$197.490_{\pm3.550}$(4.150)	$1.140_{\pm0.007}$(0.006)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.784_{\pm0.004}$(0.004)	$1.000_{\pm0.000}$	$0.859_{\pm0.003}$
RA99.0%, **TA**85.8%	0.15	$0.862_{\pm0.013}$(0.042)	$0.870_{\pm0.042}$(0.090)	$0.849_{\pm0.000}$(0.004)	$188.686_{\pm0.954}$(1.894)	$195.590_{\pm0.863}$(6.710)	$0.961_{\pm0.001}$(0.004)	$0.005_{\pm0.000}$(0.000)	$0.004_{\pm0.000}$(0.001)	$0.884_{\pm0.000}$(0.008)	$1.000_{\pm0.000}$	$0.537_{\pm0.003}$
	0.2	$0.830_{\pm0.027}$(0.043)	$0.840_{\pm0.085}$(0.020)	$0.802_{\pm0.002}$(0.003)	$187.219_{\pm0.064}$(16.168)	$194.310_{\pm0.948}$(19.830)	$0.861_{\pm0.001}$(0.002)	$0.004_{\pm0.000}$(0.000)	$0.004_{\pm0.000}$(0.001)	$0.931_{\pm0.001}$(0.005)	$1.000_{\pm0.000}$	$0.220_{\pm0.002}$
Salun	0.05	$0.997_{\pm0.003}$(0.003)	$0.993_{\pm0.012}$(0.007)	$0.949_{\pm0.001}$(0.001)	$199.599_{\pm0.207}$(0.401)	$197.440_{\pm1.244}$(2.560)	$1.980_{\pm0.050}$(0.196)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.479_{\pm0.012}$(0.053)	$1.000_{\pm0.000}$	$0.989_{\pm0.001}$
UA100%, **UA**$_{tf}$100%,	0.1	$0.975_{\pm0.019}$(0.039)	$0.927_{\pm0.023}$(0.033)	$0.899_{\pm0.001}$(0.003)	$191.973_{\pm1.616}$(0.910)	$185.220_{\pm0.918}$(8.120)	$1.169_{\pm0.002}$(0.023)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.769_{\pm0.001}$(0.019)	$1.000_{\pm0.000}$	$0.884_{\pm0.001}$
RA98.4%, **TA**86.1%	0.15	$0.961_{\pm0.022}$(0.057)	$0.860_{\pm0.040}$(0.100)	$0.850_{\pm0.001}$(0.004)	$187.825_{\pm3.461}$(1.034)	$180.307_{\pm2.908}$(8.573)	$0.969_{\pm0.002}$(0.012)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.877_{\pm0.002}$(0.015)	$1.000_{\pm0.000}$	$0.562_{\pm0.003}$
	0.2	$0.960_{\pm0.015}$(0.173)	$0.840_{\pm0.020}$(0.020)	$0.801_{\pm0.001}$(0.004)	$184.838_{\pm3.478}$(13.787)	$177.647_{\pm2.627}$(3.167)	$0.863_{\pm0.004}$(0.003)	$0.005_{\pm0.000}$(0.001)	$0.005_{\pm0.000}$(0.000)	$0.928_{\pm0.003}$(0.008)	$1.000_{\pm0.000}$	$0.230_{\pm0.009}$
SFRon	0.05	$1.000_{\pm0.000}$(0.000)	$1.000_{\pm0.000}$(0.000)	$0.948_{\pm0.001}$(0.002)	$200.000_{\pm0.000}$(0.000)	$200.000_{\pm0.000}$(0.000)	$2.264_{\pm0.254}$(0.479)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.423_{\pm0.050}$(0.110)	$1.000_{\pm0.000}$	$0.990_{\pm0.003}$
UA100%, **UA**$_{tf}$100%,	0.1	$1.000_{\pm0.000}$(0.064)	$1.000_{\pm0.000}$(0.040)	$0.900_{\pm0.001}$(0.003)	$200.000_{\pm0.000}$(7.118)	$200.000_{\pm0.000}$(6.660)	$1.266_{\pm0.044}$(0.120)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.711_{\pm0.026}$(0.077)	$1.000_{\pm0.000}$	$0.912_{\pm0.017}$
RA96.1%, **TA**84.3%	0.15	$1.000_{\pm0.000}$(0.096)	$1.000_{\pm0.000}$(0.040)	$0.850_{\pm0.002}$(0.003)	$200.000_{\pm0.000}$(13.209)	$200.000_{\pm0.000}$(11.120)	$1.009_{\pm0.012}$(0.051)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.843_{\pm0.011}$(0.049)	$1.000_{\pm0.000}$	$0.668_{\pm0.029}$
	0.2	$1.000_{\pm0.000}$(0.213)	$1.000_{\pm0.000}$(0.140)	$0.802_{\pm0.003}$(0.003)	$200.000_{\pm0.000}$(28.949)	$200.000_{\pm0.000}$(25.520)	$0.886_{\pm0.006}$(0.026)	$0.005_{\pm0.000}$(0.000)	$0.005_{\pm0.000}$(0.000)	$0.905_{\pm0.007}$(0.031)	$1.000_{\pm0.000}$	$0.358_{\pm0.017}$

Table 12: **MIACR** performance on **CIFAR-10** with **ResNet-18**.

Methods	α	**10% Forgetting**		**50% Forgetting**	
		MIACR ↑	$\hat{q}$	**MIACR** ↑	$\hat{q}$
	0.05	$0.091_{\pm 0.001}$(0.000)	$0.877_{\pm 0.004}$	$0.117_{\pm 0.010}$(0.000)	$0.899_{\pm 0.007}$
RT	0.1	$0.147_{\pm 0.000}$(0.000)	$0.589_{\pm 0.008}$	$0.201_{\pm 0.011}$(0.000)	$0.570_{\pm 0.001}$
MIA86.92% (**10% Forgetting**)	0.15	$0.203_{\pm 0.010}$(0.000)	$0.485_{\pm 0.005}$	$0.272_{\pm 0.011}$(0.000)	$0.472_{\pm 0.009}$
MIA82.79% (**50% Forgetting**)	0.2	$0.246_{\pm 0.000}$(0.000)	$0.473_{\pm 0.001}$	$0.318_{\pm 0.006}$(0.000)	$0.459_{\pm 0.003}$
	0.05	$0.039_{\pm 0.011}$(0.052)	$0.745_{\pm 0.013}$	$0.036_{\pm 0.001}$(0.081)	$0.780_{\pm 0.011}$
FT	0.1	$0.077_{\pm 0.008}$(0.070)	$0.627_{\pm 0.000}$	$0.103_{\pm 0.011}$(0.098)	$0.558_{\pm 0.012}$
MIA92.00% (**10% Forgetting**)	0.15	$0.128_{\pm 0.007}$(0.075)	$0.517_{\pm 0.008}$	$0.159_{\pm 0.011}$(0.113)	$0.494_{\pm 0.011}$
MIA92.92% (**50% Forgetting**)	0.2	$0.196_{\pm 0.003}$(0.050)	$0.483_{\pm 0.003}$	$0.244_{\pm 0.010}$(0.074)	$0.476_{\pm 0.004}$
	0.05	$0.083_{\pm 0.010}$(0.008)	$0.627_{\pm 0.011}$	$0.050_{\pm 0.016}$(0.067)	$0.547_{\pm 0.000}$
RL	0.1	$0.178_{\pm 0.027}$(0.031)	$0.572_{\pm 0.005}$	$0.137_{\pm 0.030}$(0.064)	$0.547_{\pm 0.001}$
MIA74.21% (**10% Forgetting**)	0.15	$0.272_{\pm 0.006}$(0.069)	$0.492_{\pm 0.015}$	$0.194_{\pm 0.031}$(0.078)	$0.547_{\pm 0.001}$
MIA61.15% (**50% Forgetting**)	0.2	$0.320_{\pm 0.025}$(0.074)	$0.485_{\pm 0.011}$	$0.261_{\pm 0.001}$(0.057)	$0.546_{\pm 0.000}$
	0.05	$0.012_{\pm 0.002}$(0.079)	$0.862_{\pm 0.016}$	$0.012_{\pm 0.019}$(0.105)	$0.771_{\pm 0.008}$
GA	0.1	$0.032_{\pm 0.003}$(0.115)	$0.502_{\pm 0.016}$	$0.055_{\pm 0.003}$(0.146)	$0.486_{\pm 0.005}$
MIA98.80% (**10% Forgetting**)	0.15	$0.076_{\pm 0.000}$(0.127)	$0.477_{\pm 0.007}$	$0.107_{\pm 0.016}$(0.165)	$0.474_{\pm 0.015}$
MIA98.86% (**50% Forgetting**)	0.2	$0.146_{\pm 0.016}$(0.100)	$0.476_{\pm 0.019}$	$0.164_{\pm 0.016}$(0.154)	$0.473_{\pm 0.011}$
	0.05	$0.013_{\pm 0.006}$(0.078)	$0.750_{\pm 0.014}$	$0.031_{\pm 0.003}$(0.086)	$0.635_{\pm 0.018}$
Teacher	0.1	$0.038_{\pm 0.023}$(0.109)	$0.672_{\pm 0.028}$	$0.065_{\pm 0.021}$(0.136)	$0.582_{\pm 0.013}$
MIA87.24% (**10% Forgetting**)	0.15	$0.072_{\pm 0.013}$(0.131)	$0.625_{\pm 0.029}$	$0.110_{\pm 0.017}$(0.162)	$0.548_{\pm 0.007}$
MIA93.24% (**50% Forgetting**)	0.2	$0.113_{\pm 0.008}$(0.133)	$0.588_{\pm 0.019}$	$0.159_{\pm 0.017}$(0.159)	$0.532_{\pm 0.006}$
	0.05	$0.011_{\pm 0.011}$(0.080)	$0.861_{\pm 0.012}$	$0.012_{\pm 0.002}$(0.105)	$0.748_{\pm 0.011}$
SSD	0.1	$0.031_{\pm 0.010}$(0.116)	$0.511_{\pm 0.011}$	$0.051_{\pm 0.005}$(0.150)	$0.488_{\pm 0.001}$
MIA98.78% (**10% Forgetting**)	0.15	$0.077_{\pm 0.005}$(0.126)	$0.480_{\pm 0.013}$	$0.104_{\pm 0.006}$(0.168)	$0.477_{\pm 0.015}$
MIA98.87% (**50% Forgetting**)	0.2	$0.139_{\pm 0.011}$(0.107)	$0.475_{\pm 0.013}$	$0.168_{\pm 0.012}$(0.150)	$0.477_{\pm 0.006}$
	0.05	$0.076_{\pm 0.025}$(0.015)	$0.844_{\pm 0.024}$	$0.045_{\pm 0.008}$(0.072)	$0.863_{\pm 0.025}$
NegGrad+	0.1	$0.128_{\pm 0.018}$(0.019)	$0.481_{\pm 0.009}$	$0.109_{\pm 0.007}$(0.092)	$0.511_{\pm 0.008}$
MIA90.30% (**10% Forgetting**)	0.15	$0.174_{\pm 0.022}$(0.029)	$0.480_{\pm 0.005}$	$0.167_{\pm 0.017}$(0.105)	$0.477_{\pm 0.010}$
MIA93.82% (**50% Forgetting**)	0.2	$0.213_{\pm 0.012}$(0.033)	$0.480_{\pm 0.004}$	$0.230_{\pm 0.014}$(0.088)	$0.472_{\pm 0.008}$
	0.05	$0.055_{\pm 0.014}$(0.036)	$0.691_{\pm 0.011}$	$0.044_{\pm 0.001}$(0.073)	$0.670_{\pm 0.008}$
Salun	0.1	$0.113_{\pm 0.009}$(0.034)	$0.681_{\pm 0.013}$	$0.115_{\pm 0.009}$(0.086)	$0.630_{\pm 0.009}$
MIA57.58% (**10% Forgetting**)	0.15	$0.198_{\pm 0.006}$(0.005)	$0.642_{\pm 0.015}$	$0.170_{\pm 0.009}$(0.102)	$0.610_{\pm 0.003}$
MIA59.12% (**50% Forgetting**)	0.2	$0.267_{\pm 0.009}$(0.021)	$0.608_{\pm 0.011}$	$0.220_{\pm 0.005}$(0.098)	$0.586_{\pm 0.005}$
	0.05	$0.017_{\pm 0.001}$(0.074)	$0.711_{\pm 0.009}$	$0.017_{\pm 0.002}$(0.100)	$0.715_{\pm 0.008}$
SFRon	0.1	$0.040_{\pm 0.004}$(0.107)	$0.626_{\pm 0.025}$	$0.046_{\pm 0.002}$(0.155)	$0.562_{\pm 0.013}$
MIA91.55% (**10% Forgetting**)	0.15	$0.113_{\pm 0.003}$(0.090)	$0.517_{\pm 0.003}$	$0.134_{\pm 0.013}$(0.138)	$0.498_{\pm 0.003}$
MIA92.52% (**50% Forgetting**)	0.2	$0.184_{\pm 0.002}$(0.062)	$0.487_{\pm 0.002}$	$0.206_{\pm 0.014}$(0.112)	$0.483_{\pm 0.002}$