Highlights

Does Audio Augmentation Aid Audio Unlearning? A Comprehensive Evaluation Across Methods

Moulik Gupta, Achyut Mani Tripathi

- We present the first systematic empirical analysis that evaluates the impact of different audio augmentation techniques on audio unlearning performance across three major audio architectures (AST, Audio-Mamba, ResNet-18) and four benchmark datasets.
- We thoroughly assess nine prominent machine unlearning approaches, namely Amnesiac, Bad Teacher, Boundary, Fisher, Gradient Ascent, Retrain, SSD, UNSIR, and SCRUB, demonstrating how augmentation-guided audio unlearning significantly reduces the performance gap between retrained and unlearned audio models while improving class-wise forgetting.
- Through extensive experiments, we show that carefully chosen audio augmentations (especially time masking and additive noise) consistently enhance audio unlearning stability and effectiveness, highlighting audio augmentation as an essential component for building robust, reliable, and privacy-preserving audio unlearning systems.

Does Audio Augmentation Aid Audio Unlearning? A Comprehensive Evaluation Across Methods

Moulik Gupta, Achyut Mani Tripathi[a]

[a] *Department of Computer Science & Engineering, Indian Institute of Technology, Dharwad, 580007, Karnataka, India*
[b] *Department of Electrical Electronics & Engineering, Guru Gobind Singh Indraprastha University, New Delhi, 221005, Delhi, India*

Abstract

Machine unlearning aims to remove the influence of specific training samples from a learned model while preserving its utility on the retained data. In the audio domain, this objective becomes more challenging due to the inherent temporal structure and sequential complexity of audio signals, as well as the strong memorization tendencies of deep audio models. Although audio data augmentation is known to influence memorization, its contribution to machine unlearning has not been systematically examined. In this work, we present the first comprehensive study that evaluates nine state-of-the-art machine unlearning methods, Viz. Amnesiac, Bad Teacher, Boundary, Fisher, Gradient Ascent, Retrain, SSD, UNSIR, and SCRUB across three prominent audio architectures: AST, Audio-Mamba, and ResNet-18. Using four benchmark audio datasets (Google Speech Commands V1, ESC-10, GTZAN, and US8K), we investigate how incorporating audio augmentation affects the effectiveness of machine unlearning for each model. Our results demonstrate that audio augmentations, particularly time masking and additive noise, consistently improve unlearning performance and substantially reduce the gap between retrained and unlearned models. This study demonstrates that carefully integrating audio augmentations is crucial for enhancing privacy-preserving unlearning and producing more reliable and robust audio models.

Keywords: Audio classification, Audio state space model, audio Unlearning, Residual network, State space model, Spectrogram.

1. Introduction

Large datasets are essential for training machine learning models, but they often contain sensitive, copyrighted, or private information. In some cases, such data may even be collected without user consent. Regulations such as the General Data Protection Regulation (GDPR) [1] mandate the removal of this information, yet retraining models from scratch is often costly and impractical. Machine unlearning [2] addresses this challenge by enabling the removal of specific data from a trained model without full retraining, ensuring that the model behaves as if the deleted data were never part of its training set. Machine unlearning has been widely explored in the domain of image [3] classification. In general, unlearning techniques can be categorized as exact or approximate [4] methods. Exact methods aim to completely remove targeted data by selectively retraining parts of the model, though they remain computationally expensive. Approximate methods, on the other hand, reduce the influence of the data without guaranteeing full elimination, making them more efficient and better suited for frequent or large-scale unlearning scenarios. Several strategies have been proposed to weaken the influence of forgotten data. For example, Golatkar et al. [5] introduced a perturbation-based function that adjusts model weights to erase information about the target data. Influence-based methods [6] rely on one-shot updates to enforce unlearning of specific samples. Gradient ascent has also been extensively studied, particularly in the context of large language models [7]. Other notable approaches include logits-based unlearning [8], and mitigation of highly memorized samples [9].

Although numerous approaches have been proposed to evaluate machine unlearning in image classification, the domains of audio and speech processing have received far less attention. Only a limited number of studies address audio unlearning. For instance, [10] introduced the first benchmark study, evaluating eight machine unlearning methods by unlearning Wave2Vec and HuBERT models for spoken language understanding. Likewise, Cheng et al. [11] investigated several unlearning methods for speech tasks, revealing why unlearning in speech is inherently more difficult compared to image or text data. Label smoothing [12] and network pruning [13] are widely adopted machine unlearning techniques designed to reduce model memorization and complexity. Zhao et al. [14] showed that the presence of memorized information substantially increases the difficulty of unlearning. Although methods mentioned above have opened promising directions in the literature, the role of audio data augmentation in audio unlearning has remained largely unexplored. Despite the critical role of audio data augmentation in controlling memorization and mitigating bias in audio machine learning [15], its systematic impact on audio unlearning remains largely unexplored, representing a significant research gap.

To bridge this gap, this work presents the first comprehensive empirical evaluation of the impact of various audio data augmentation strategies on unlearning performance. Our study spans multiple state-of-the-art class-wise unlearning methods and three diverse audio datasets, providing clear evidence that carefully chosen audio augmentation techniques can significantly enhance the effectiveness of audio unlearning. Our key contributions are as follows: we provide a comprehensive evaluation of five state-of-the-art machine unlearning methods in the audio domain across three benchmark datasets, complemented by a systematic study of audio data augmentation. Our results demonstrate that appropriate augmentations substantially reduce the performance gap between unlearned audio models and those retrained from scratch, while also enhancing unlearning efficiency. Additionally, through a detailed class-wise forgetting analysis, we reveal the sensitivity of unlearning methods to augmentation strategies, highlighting their pivotal role in building robust and reliable audio unlearning frameworks.

2. Problem Setup

Assume $D = \{x_j, y_j\}_{j=1}^{N}$ is a training dataset containing N audio samples, where each sample is associated with a class label $y \in 1 \ldots C$, and C represents the total number of classes.

A machine unlearning algorithm seeks to eliminate the influence of a specific subset of the audio dataset D, represented as $D_F \in D$, from the previously trained model f_{θ_0}, which was trained on D with learned parameters θ_0. The traditional retraining approach involves training a model f_{θ_r} from scratch on the remaining audio dataset $D_R \in \{D - D_F\}$ to obtain the updated set of weights θ_r. The model f_{θ_r} is trained without D_F and is therefore free from the influence of audio samples in D_F. However, this process becomes computationally expensive as the size of the audio dataset increases. To address the above limitation, machine unlearning methods aim to train an unlearned model f_{θ_u}, where the initial set of weights θ_u is set to the previously learned parameters, i.e., $\theta_u = \theta_0$. The machine unlearning procedure then updates f_{θ_u} in the presence of both D_F and D_R such

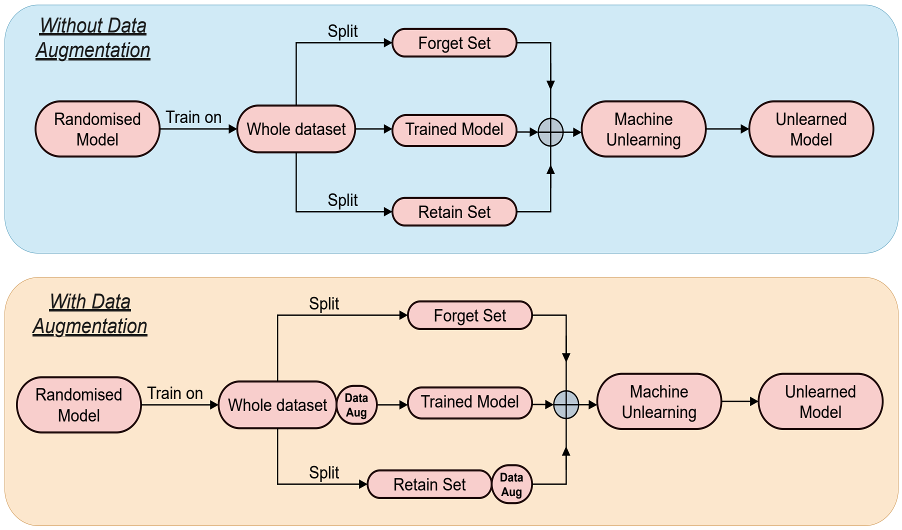


Figure 1: Overall workflow of the methodology used to evaluate the performance of various unlearning methods under different data augmentation settings

that the performance of f_{θ_u} closely matches that of f_{θ_r} when evaluated on D_F and D_R. Figure 1 illustrates the overall workflow of the methodology adopted to assess the performance of different unlearning methods under diverse data augmentation settings. The performance of the unlearned model is evaluated by computing the metric gap ($MG = MG_r - MG_u$) between f_{θ_r} and f_{θ_u} under different data augmentation settings. The objective of the machine unleanring algorithm under different data augmentation settings is to minimize this metric gap so that its value approaches zero. In this work, our primary aim is to explore whether incorporating audio data augmentation techniques can further reduce the metric gap. The details of the various metrics used to compute the metric gaps are provided in Section 3.

3. Evaluated Models

To study the impact of audio data augmentation methods, we selected an audio classification task using three widely adopted audio models: the Audio Spectrogram Transformer (AST)[16], the Audio State Space Model (Audio Mamba)[17], and a Residual Network (ResNet) [18]. We included the ResNet model to ensure a fair comparison between models that divide the input spectrogram into patches (AST and Audio Mamba) and those that operate directly on the full spectrogram without patching (ResNet). All three models are first fine-tuned on their respective audio classification datasets and subsequently unlearned both in the presence and absence of audio data augmentation methods to evaluate the resulting metric gaps.

4. Experiments

4.1. Datasets

We conducted experiments on four audio datasets (Table 1), splitting the audio dataset D into retained D_R and forget D_F sets, where D_f contains samples from classes to be removed from the model.

Table 1: Dataset Specifications

Dataset	Sampling Rate	Classes	#Samples
GSCD V1 [19]	16 kHz	10	23,682
ESC-10 [20]	44.1 kHz	10	400
US8K [21]	16 kHz	10	8,732
GTZAN [22]	22.05 kHz	10	1000

4.2. Implementation Details

The experiments were conducted using Python 3.10.16 on Ubuntu 24.04.1 LTS with CUDA 12.1, running on a single NVIDIA RTX 4060 GPU. Model training utilized the AdamW optimizer with a learning rate of $2e^{-5}$, a batch size of 16, and early stopping with a patience of 7. Cross-entropy loss was used during training. The forgetting rate was fixed at 10% for all experiments. Initially, all three models were fine-tuned on the four datasets, both with and without data augmentation. Afterward, a target class (class 0 for all datasets in this study) was randomly selected, and experiments were carried out to unlearn all audio samples associated with that class. During spectrogram extraction, the number of Mel filters, FFT size, and window overlap were set to 128, 1024, and 512, respectively. The size of the spectrogram is $128 \times 100t \times 1$ while training the AST and Audio Mamba model. Here t indicates the length of the signal in seconds. The value of patch size is set to 16 while training the AST and Audio Mamba model. The value of the learning rate is set to $2e^{-4}$ while training the three models.

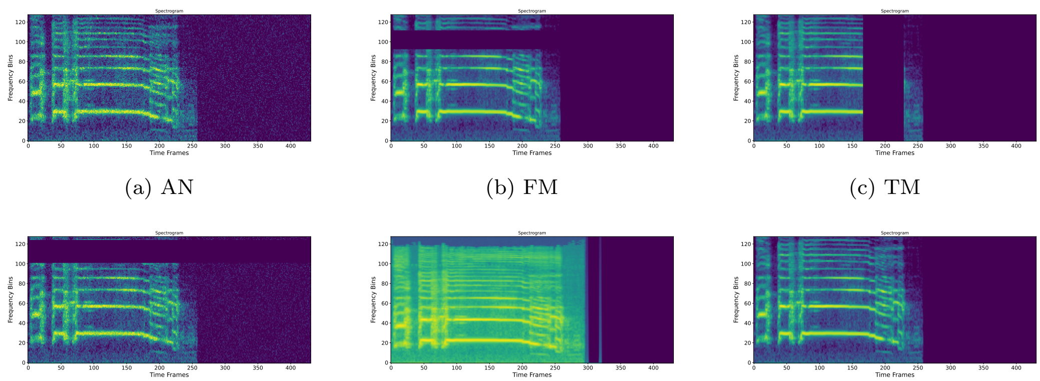
4.3. Evaluation Metrics

We evaluate the model using unlearning accuracy (UNA, lower better), remaining accuracy (RMA, higher better), testing accuracy (TEA, higher better), and membership inference attack (MIA, lower better). Additionally, we introduce two metrics: average gap (AGP, lower better), which quantifies the mean difference across these four metrics between retrained and unlearned models, and runtime efficiency (RTE, lower better), which measures unlearning time (or retraining time for the retained dataset).

- **UNA**: It is defined as the accuracy achieved by the unlearned model on the forget dataset D_F.
- **RMA**: It is defined as the accuracy of the unlearned model on the remaining dataset D_R.
- **TEA**: It is defined as the accuracy of the unlearned model on the test dataset D_{test}.
- **MIA**: This metric evaluates how susceptible the unlearned model is to membership inference attacks.
- **AGP**: It is calculated as the average absolute difference between the UNA, TEA, and MIA metrics of the unlearned model and those of the retrained model. This metric indicates how closely the unlearned model aligns with the retrained model.
- **RTE**: It measures the execution time required by each unlearning method. In the case of retraining, it corresponds to the time taken to retrain on the remaining dataset.

4.4. Data Augmentations

To investigate the effect of data augmentation on unlearning performance, we applied various audio data augmentation techniques [23] individually and in combination. Specifically, we employed five augmentation strategies: Time Masking (TM), Frequency Masking (FM), Additive Noise (AN), Time Stretch (TS), and a combined method of AN and FM. The abbreviation NA denotes the condition where no augmentation is applied during the unlearning of an audio model. Figure 2 depicts the spectrograms corresponding to the audio signal after applying different data augmentation methods.



(a) AN (b) FM (c) TM

(d) AN+FM (e) TS (f) NA

Figure 2: Visualization of Spectrogram Obtained Under Different Audio Data Augmentation Settings

4.5. Unlearning Methods

The class-wise unlearning performance of the three audio models is evaluated using five state-of-the-art approaches, Viz. Amnesiac [24], Bad Teacher [25], Boundary [26], Fisher [27], and Scrub [28]. For fairness, the performance of these unlearned models is compared against the audio model retrained on the remaining data samples after excluding the target class.

5. Results

5.1. Performance of Audio Data Augmentation Methods for Unlearning the AST Model

1. Combined perturbation AN+FM frequently results in the poorest performance. AN+FM appears repeatedly as the Worst AGP, Worst MIA, or Worst UNA across multiple methods and datasets, indicating that jointly applying noise and frequency masking severely disrupts AST representations.
2. Temporal Shift (TS) consistently emerges as a strong augmentation for unlearning stability. TS is often listed as Best AGP or Best MIA in methods such as Boundary, SCRUB, and Gradient Ascent, suggesting that temporal modifications help AST effectively decouple target samples during unlearning.
3. Noise Augmentation (AN) generally improves MIA robustness. AN frequently appears as the Best MIA across datasets, indicating that

injecting noise reduces latent separability and mitigates membership inference vulnerabilities.

4. Frequency Masking (FM) shows dataset-dependent behavior. FM sometimes yields strong performance (e.g., best AGP for BadTeacher, Fisher, Gradient Ascent), but in other cases appears in Worst AGP/MIA/UNA categories, highlighting its inconsistent impact across datasets.
5. Boundary and Gradient Ascent display high sensitivity to augmentation choice. These methods show large performance fluctuations depending on augmentation, with certain augmentations (TS, FM) enhancing unlearning and others (NA, TM) significantly degrading it.
6. Mask-based augmentations tend to preserve retained-class accuracy (UNA). Augmentations such as AN, FM, and TS frequently appear under Best UNA, suggesting they help maintain performance on retained data while forgetting the targeted samples.
7. Time Mask (TM) often correlates with poor forgetting metrics. TM repeatedly appears in Worst AGP or Worst MIA (e.g., Fisher, BadTeacher, Gradient Ascent), implying that temporal occlusion harms AST's ability to unlearn effectively.
8. UNSIR exhibits strong robustness across augmentation types. UNSIR consistently reports *ALL* for Best UNA and maintains competitive MIA scores, indicating that it is less sensitive to augmentation tuning compared to other methods.
9. The TS, FM, and AN generally strengthen unlearning metrics, whereas TM and AN+FM consistently degrade performance. This indicates that individual temporal or spectral perturbations are beneficial, but combined or severe masking has an adverse effect on AST unlearning behavior.

5.2. ***Performance of Audio Data Augmentation Methods for Unlearning the Audio Mamba Model***

1. AN+FM consistently undermines unlearning performance in AudioMamba. AN+FM frequently appears as the Worst AGP or Worst MIA across multiple datasets (ESC10, GTZAN, GSCD, US8K). This suggests that the joint application of additive noise and frequency masking significantly disrupts the internal state-space representations of AudioMamba and destabilizes the forgetting process.

Table 2: **Evaluation of Machine Unlearning Methods on the ESC-10 and GSCD Datasets with the AST Model**

Datasets		ESC10						GSCD					
Unlearning Methods	**Aug**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**
Amnesiac	NA	0	100	93.06	16.67	0.69	44.93	0	99.97	94.09	27.52	5.66	756.11
	TM	0	100	95.14	70.83	15.74	77.9	0	98.95	93.45	25.46	8.26	551.73
	FM	0	100	95.83	79.17	23.84	73.41	0	98.29	93.51	29.75	0.88	617.35
	AN	0	100	84.03	54.17	20.83	58.2	0	99.36	91.98	34.69	4.2	618.23
	AN+FM	0	100	94.44	79.17	25.69	88.21	0	97.84	94.48	48.97	2.3	552.97
	TS	0	100	90.97	12.5	0.93	269.02	0	99.47	93.23	11.02	7.96	876.38
Bad Teacher	NA	3.12	99.11	90.28	0	8.22	29.43	4.26	98.52	94.09	0	16.25	593.98
	TM	3.12	100	92.36	0	9.84	29.59	1.02	96.92	92.73	0	17.33	606.75
	FM	3.12	98.21	90.28	0	5.9	29.57	2.21	96.19	93.19	0	11.64	613.09
	AN	3.12	100	80.56	0	7.75	30.21	2.96	96.64	91.38	0	16.96	608.6
	AN+FM	12.5	92.41	81.94	0	9.03	29.95	1.83	96.08	94.22	0	15.05	617.11
	TS	18.75	98.66	88.19	0	12.27	154.56	5.66	97.41	92.76	0	13.68	1791.4
Boundary	NA	75	95.98	83.33	75	48.38	5.82	61.85	96.66	90.84	93.11	37.91	106.51
	TM	31.25	59.82	59.03	87.5	42.82	5.82	42.03	96.83	92.47	78.45	23.93	105.95
	FM	59.38	90.62	79.86	70.83	46.18	5.81	56.09	95.13	92.28	95.87	41.23	105.24
	AN	68.75	100	90.28	20.83	30.56	5.83	58.94	94.47	90.84	93.32	36.65	105.41
	AN+FM	40.62	83.48	72.92	75	43.63	5.21	50.54	95.16	93.79	4.07	30.07	104.92
	TS	0	60.71	52.08	45.83	25	22.89	37.28	90.29	85.45	39.25	16.9	239.54
Fisher	NA	9.38	99.55	92.36	0	9.61	64.26	1.13	97.61	91.77	2.01	14.82	5527.26
	TM	15.62	98.66	90.28	0	14.7	58.8	0.65	98.07	93.38	2.17	16.26	5374.56
	FM	9.38	98.21	93.06	0	7.06	59.86	0.16	96.14	93.36	0.43	10.76	5554.2
	AN	0	98.66	94.44	0	2.08	63.46	0.54	96.01	91.22	0.65	15.98	5897.54
	AN+FM	0	99.11	94.44	0	2.08	64.96	0.16	96.4	94.16	0.54	14.33	5908.66
	TS	0	98.21	89.58	0	5.56	89.82	0.48	98.46	93.06	0.22	11.78	5568.75
Gradient Ascent	NA	100	100	93.75	100	61.57	14.77	0.11	83.42	77.21	93.11	21.87	217.17
	TM	0	66.52	66.67	83.33	28.47	14.99	0	84.97	81.29	95.44	19.32	226.64
	FM	100	99.11	95.83	20.83	37.73	15.26	99.08	97.57	93.62	91.15	53.54	230.45
	AN	0	65.62	38.19	83.33	45.83	15.66	97.36	98.41	94.11	80.78	44.32	230.08
	AN+FM	0	93.3	79.86	25	11.11	15.22	99.52	97.77	94.44	91.53	49.67	242.39
	TS	0	96.43	80.56	66.67	22.45	295.32	97.04	99.21	92.72	77.47	47.14	2297.97
Retrain	NA	0	100	95.14	16.67	0	170.62	0	99.24	93.36	43.76	0	743.2
	TM	0	100	93.75	25	0	162.45	0	99.72	93.75	49.95	0	705.79
	FM	0	100	96.53	8.33	0	154.33	0	99.6	94.96	30.94	0	705.28
	AN	0	100	96.53	4.17	0	145.96	0	99.59	93.9	45.39	0	715.58
	AN+FM	0	99.11	92.36	4.17	0	80.26	0	97.91	94.97	42.56	0	876.98
	TS	0	100	93.75	12.5	0	984.61	0	99.72	93.88	34.26	0	4609.62
SCRUB	NA	0	78.12	65.97	66.67	26.39	30.08	99.84	99.82	94.09	96.63	51.15	585.72
	TM	0	96.43	87.5	41.67	7.64	29.06	99.95	99.87	95.02	97.34	49.54	584.23
	FM	0	100	92.36	25	6.94	29.15	99.08	98.73	94.7	92.94	53.78	586.66
	AN	0	99.55	93.75	41.67	13.43	29.09	98.22	99.08	94.7	92.29	48.64	582.44
	AN+FM	0	99.55	95.14	20.83	6.48	29.12	99.89	98.65	94.96	90.1	49.14	585.55
	TS	0	99.55	95.83	20.83	3.47	171.86	99.41	99.82	94.53	94.3	53.37	1810.26
SSD	NA	0	13.39	13.19	37.5	34.26	1.71	0	98.89	93.28	9.83	11.34	49.46
	TM	0	22.32	22.22	83.33	43.29	3.11	0	48.98	49.2	6.84	29.22	51.72
	FM	0	98.21	92.36	0	4.17	2.95	98.65	97.73	93.75	75.79	48.23	51.11
	AN	0	10.71	11.11	0	29.86	2.97	55.28	98.52	94.18	0.05	33.63	50.63
	AN+FM	0	42.86	48.61	95.83	45.14	2.79	99.73	97.56	94.27	93.81	50.56	50.3
	TS	0	11.61	11.11	37.5	35.88	22.35	97.04	99.49	93.19	32.84	33.05	184.89
UNSIR	NA	0	80.36	74.31	83.33	29.17	1040.9	0	89.08	84.07	25.35	9.23	254.96
	TM	0	91.96	73.61	54.17	16.44	1139.53	0	88.89	85.92	24.1	11.22	264.29
	FM	0	79.46	88.89	87.5	28.94	1000.06	0	84.83	84.37	68.19	15.94	264.75
	AN	0	72.77	65.28	33.33	20.14	1085.82	0	86.82	83.42	34.15	7.24	265.35
	AN+FM	0	64.73	62.5	58.33	28.01	1119.4	0	60.58	65.02	83.66	23.68	277.41
	TS	0	85.71	76.39	58.33	21.06	1129.28	0	84.91	81.4	19.6	9.05	1514.39

Table 3: Evaluation of Machine Unlearning Methods on the Us8K and GTZAN Datasets with the AST Model

Datasets		US8K						GTZAN					
Unlearning Methods	**Aug**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**
Amnesiac	NA	0	99.97	84.84	3.88	5.91	164.85	0	100	78.12	38.46	10.52	8563.75
	TM	0	99.51	88.3	5.25	11.3	162.49	0	100	80.21	60.26	18.38	14140.56
	FM	0	96.63	88.3	8.75	12.67	162.04	0	99.86	80.21	57.69	16.45	8500.63
	AN	0	99.98	88.3	7.25	12.95	185.76	0	100	79.17	21.79	5.98	7323.21
	AN+FM	0	97.96	89.49	13.62	7.6	161.4	0	99.86	80.21	37.18	10.44	7151.91
	TS	0	99.13	84.04	11.12	7.32	163.74	0	100	79.17	39.13	9.12	9246.11
Bad Teacher	NA	0	97.34	86.3	0	6.72	176.98	0	98.76	76.04	0	4.38	8359.77
	TM	0	96.18	84.97	0	12.47	181.59	0	99.18	75	0	3.45	8426.67
	FM	0	93.83	87.23	0	15.68	178.66	0	92.99	70.83	0	5.9	8790.25
	AN	0	96.81	82.71	0	15.99	183.1	0	99.59	70.83	0	4.06	8298.18
	AN+FM	0	94.18	85.51	0	12.94	176.17	1.25	97.39	73.96	0	4.45	7498.65
	TS	0	93.69	82.31	0	10.63	167.7	0	98.78	73.96	0	8.44	6041.4
Boundary	NA	4.5	97.84	87.77	19.62	1.81	37.17	16.67	66.35	50	48.72	28.87	1717.83
	TM	1.62	93.1	83.11	11.25	9.89	37.42	33.75	81.04	60.42	34.62	27.68	1683.64
	FM	13.5	93.81	90.03	9.88	17.38	36.07	38.75	84.62	63.54	34.62	21.68	1647.57
	AN	15.75	96.26	83.24	33.62	9.86	37.35	35.83	73.21	50	37.18	32.78	1491.95
	AN+FM	20.12	94.97	89.49	10.12	15.47	37.36	50	81.87	63.54	38.46	31.01	1551.04
	TS	44.5	89.5	78.32	14.12	22.09	37.67	27.22	56.11	40.97	53.62	35.75	1424.93
Fisher	NA	11.62	94.62	84.57	1	10.84	2062.39	20.42	89.97	69.79	2.56	12.41	372.42
	TM	12.75	92.94	82.71	0.75	17.23	2061.69	10	85.71	67.71	2.56	8.35	372.68
	FM	2	91.19	87.5	2.12	15.55	1825.13	15	93.68	79.17	1.28	12.56	371.83
	AN	2.25	96.28	85.24	0.12	15.86	1916.67	8.75	87.64	71.88	1.28	6.2	371.58
	AN+FM	1.5	91.44	87.9	0.12	12.6	1806.02	5	93.68	72.92	0	6.05	372.13
	TS	12.12	90.27	80.19	0.88	15.09	1849.7	23.06	92.53	78.12	0	14.74	376.86
Gradient Ascent	NA	0	55.03	49.2	98.12	38.71	64.14	100	100	75	85.9	60.71	712.6
	TM	0	56.62	48.14	20.5	17.92	65.29	100	100	76.04	87.18	62.07	675.76
	FM	0	65.16	59.84	95.5	25.73	67.58	100	98.9	78.12	76.92	55.5	689.07
	AN	0	62.51	56.78	95	26.8	70.25	100	100	73.96	85.9	62.42	660.36
	AN+FM	99.12	96.72	87.37	88.5	51.11	76.21	100	99.31	76.04	92.31	61.46	660.8
	TS	0	56.86	57.31	88.38	27.34	64.68	100	100	79.17	84.06	57.43	658.85
Retrain	NA	0	98.62	86.84	19.62	0	201.29	0	100	80.21	8.97	0	7176.8
	TM	0	99.4	85.77	36.62	0	206.12	0	100	80.21	5.13	0	8183.44
	FM	0	98.78	87.9	46.38	0	353.59	0	94.37	71.88	16.67	0	9814.65
	AN	0	98.79	86.44	44.25	0	226.43	0	100	79.17	3.85	0	8668.53
	AN+FM	0	98.25	88.7	35.62	0	284.39	0	99.31	77.08	8.97	0	11843.44
	TS	0	100	82.58	31.62	0	220.24	0	100	83.33	15.94	0	7339.6
SCRUB	NA	94.25	99.83	88.7	63.75	46.75	182.37	90.83	100	77.08	47.44	44.14	9982.75
	TM	94.38	99.17	89.76	64.38	42.04	184.19	95	100	76.04	53.85	49.29	10389.21
	FM	86.88	97.61	89.63	52.38	31.53	185.37	100	99.86	79.17	58.97	49.87	10599.07
	AN	96.5	99.9	89.1	74.25	43.05	182.88	100	100	73.96	97.44	66.27	11629.8
	AN+FM	96.38	98.46	88.83	77	45.96	184.73	98.75	99.86	75	87.18	59.68	10499.8
	TS	86	99.75	84.31	51.25	35.78	184.09	85.28	100	85.42	42.03	37.82	10422.64
SSD	NA	0	12.82	13.96	0	30.83	16.86	0	99.45	78.12	0	3.69	533.19
	TM	0.5	97.2	88.96	1.38	12.98	16.36	0	11.13	8.33	35.9	34.21	529.78
	FM	87.88	95.5	88.03	0.38	44.67	17.51	0	99.04	81.25	0	8.68	529.76
	AN	96	99.27	86.3	5.12	45.09	16.91	0	47.12	41.67	10.26	14.64	532.84
	AN+FM	99.38	96.61	87.5	91.88	52.27	18.13	100	99.31	78.12	60.26	50.77	548.48
	TS	97.12	95.47	80.59	76.5	48	17.2	0	95.52	71.88	0	9.13	553.62
UNSIR	NA	0	88.56	71.41	37.88	11.23	105.83	0	67.72	51.04	39.74	19.98	6171.84
	TM	0	87.37	76.6	41.5	4.68	107.43	0	66.48	44.79	32.05	20.78	6106.65
	FM	0	80.12	72.87	36.12	8.43	107.43	0	54.81	41.67	38.46	17.33	6165.84
	AN	0	89.82	76.99	33.62	6.69	108.65	0	71.7	51.04	34.62	19.63	5965.1
	AN+FM	0	82.49	71.01	39.25	7.1	111.83	0	65.8	46.88	37.18	19.47	5982.98
	TS	0	87.43	72.61	33	3.78	104.34	0	63.04	47.22	24.64	14.94	5499.11

2. Time Mask (TM) emerges as a powerful augmentation for Mamba-based architectures. TM appears repeatedly as Best AGP, Best MIA, or Best UNA for several methods (BadTeacher, Fisher, Gradient Ascent, UNSIR) across GTZAN, GSCD, and US8K, indicating that temporal occlusion interacts favorably with the Mamba model's sequential state-update mechanism.
3. Noise Augmentation (AN) strengthens forgetting and MIA robustness across many methods. AN frequently appears as Best AGP or Best MIA in ESC10, GTZAN, and US8K. This highlights that additive noise helps Audio-Mamba erase specific memory traces while improving privacy robustness during the unlearning process.
4. TS (Temporal Shift) provides consistent improvements in certain datasets but is unstable in others. TS is the Best AGP or Best MIA for several methods (SCRUB, UNSIR, Boundary) in ESC10 and GSCD. However, it is also often the Worst AGP or Worst UNA in GTZAN and US8K. This indicates strong dataset-dependent behavior of TS within the Mamba architecture.
5. FM (Frequency Masking) has highly variable performance depending on the dataset. FM appears as the Best AGP or Best MIA for Boundary, Fisher, and SSD on GTZAN. Yet in ESC10 and US8K, FM is frequently the Worst AGP or Worst MIA. These inconsistencies reflect how Audio-Mamba's frequency-sensitive components react differently across dataset characteristics.
6. Gradient Ascent and BadTeacher show high sensitivity to augmentation selection. These two methods show dramatic shifts in performance based on augmentation. For example, Gradient Ascent benefits strongly from TM or AN in GSCD and US8K but performs poorly with TS or AN+FM. BadTeacher also swings widely, with AN or TM performing well while AN+FM and FM perform poorly in several datasets.
7. Mask-based augmentations (TM, FM) help maintain UNA in many methods. TM and FM frequently appear under Best UNA, especially for BadTeacher, UNSIR, Boundary, and SSD. These augmentations appear to preserve retained-class accuracy, possibly by regularizing temporal-frequency patterns rather than disrupting full representations.
8. UNSIR consistently demonstrates strong, stable unlearning across datasets. UNSIR often reports Best AGP, Best MIA, or Best UNA for ESC10, GTZAN, GSCD, and US8K. It rarely appears in the worst categories,

indicating that its structured optimization strategy remains stable under varying augmentations.

9. The TM, NA, and FM (individually) frequently enhance unlearning metrics, whereas AN+FM and occasionally TS degrade performance. This suggests that Audio-Mamba benefits from single-axis perturbations (temporal or spectral), while combined aggressive distortions disrupt its sequential modeling behavior and lead to poor unlearning outcomes.

*5.3. **Performance of Audio Data Augmentation Methods for Unlearning the ResNet-18 Model***

1. AN+FM augmentation is frequently associated with the worst performance across methods and datasets. AN+FM repeatedly appears as the Worst AGP, Worst MIA, or Worst UNA, highlighting that the combination of additive noise and frequency masking destabilizes the ResNet-18 features and harms the effectiveness of unlearning.
2. Temporal Shift (TS) is a robust and often top-performing augmentation for ResNet-18 unlearning. TS appears consistently as Best AGP or Best MIA for multiple methods (Boundary, UNSIR, Fisher, Gradient Ascent) and datasets, suggesting that temporal displacement aids the network in decoupling target sample information during forgetting.
3. Noise Augmentation (AN) strengthens MIA resistance in many scenarios. AN frequently shows up as Best MIA across datasets (ESC10, GTZAN, GSCD, US8K), confirming that noise reduces sample distinctiveness and improves privacy characteristics during unlearning.
4. Frequency Masking (FM) exhibits significant variance in performance across different datasets. FM sometimes appears as a best-case augmentation (e.g., Boundary for GSCD; Fisher for ESC10), while in other cases it is consistently the worst performer (GTZAN, US8K), emphasizing dataset-specific sensitivity.
5. Gradient Ascent and Boundary methods are highly sensitive to augmentation choice. Both methods display wide fluctuations: TS and FM often improve AGP or MIA, while TM, AN, and especially AN+FM frequently degrade results. This suggests that gradient-based forgetting is unstable and heavily influenced by the augmentation profile.

Table 4: Evaluation of Machine Unlearning Methods on the ESC-10 and GSCD Datasets with the Audio Mamba Model

Datasets		**ESC10**						**GSCD**					
Unlearning Methods	**Aug**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**
Amnesiac	NA	0	100	91.67	20.83	0.23	24.13	0	99.49	89.61	4.61	10.54	951.94
	TM	0	99.55	91.67	45.83	8.56	38.15	0	98.67	89.74	21.39	3.94	1181.63
	FM	0	100	88.89	62.5	18.75	38.24	0	98.56	90.95	9.93	10.14	913.75
	AN	0	100	89.58	45.83	13.66	33.95	0	99.38	87.99	15.09	9.52	1056.68
	AN+FM	0	100	91.67	100	28.24	63.67	0	95.31	90.1	6.79	12.37	928.88
	TS	0	100	90.97	37.5	8.1	325.58	0	97.52	89.33	10.42	9.2	1357.28
Bad Teacher	NA	0	99.55	86.11	0	9.03	21.62	0	95.59	88.91	0	11.84	1011.8
	TM	0	99.11	90.28	0	7.18	21.78	0	95.29	88.15	0	11.6	1024.81
	FM	3.12	100	91.67	0	6.37	21.67	0.11	95.5	91.77	0	13.49	998.07
	AN	0	99.55	90.28	0	3.7	22.07	0	96.99	89.05	0	14.28	981.74
	AN+FM	3.12	98.66	88.19	0	7.29	22.2	0	93.25	90.3	0	14.56	1012.72
	TS	0	99.55	91.67	0	6.48	152.76	0.11	93.47	89.93	0	12.51	2350.33
Boundary	NA	46.88	29.02	20.83	0	46.41	7.19	66.16	69.2	64.05	83.12	46.2	219.62
	TM	56.25	11.16	9.72	100	72.22	7.16	45.42	69.67	66.66	17.16	28.19	221.98
	FM	59.38	43.75	45.14	0	40.62	7.21	36.58	67.07	67.22	20.68	26.66	223.49
	AN	90.62	83.93	67.36	100	69.33	7.22	55.28	48.19	46.93	11.4	42.87	222.36
	AN+FM	84.38	42.86	38.89	0	50.81	7.3	30.55	58.3	61.06	13.68	29.93	221.31
	TS	59.38	30.36	24.31	12.5	44.56	26.92	12.93	64.86	62.03	53.64	19.75	368.36
Fisher	NA	31.25	38.84	42.36	37.5	32.64	195.06	0.05	94.42	86.96	1.68	11.68	14280.95
	TM	15.62	40.62	38.89	12.5	25.35	197.21	0	92.54	86.76	3.2	11	14785.72
	FM	12.5	62.05	53.47	4.17	20.83	193.64	0.11	93.24	88.15	2.77	13.5	15105.13
	AN	3.12	54.91	42.36	4.17	19.33	196.87	0.16	93.61	86.22	3.85	13.91	15472.7
	AN+FM	3.12	89.73	72.92	0	12.38	195.5	0	89.69	86.16	5.37	14.15	15351.06
	TS	0	29.46	21.53	4.17	28.47	221.28	2.48	87.88	82.6	3.42	14.6	15104.33
Gradient Ascent	NA	100	100	93.06	45.83	41.9	12.32	0	11.29	10.73	0	37.63	373.46
	TM	100	100	90.28	54.17	44.68	12.39	0	11.14	10.82	0	37.38	392.37
	FM	3.12	13.84	13.89	58.33	43.4	12.43	0	11.1	10.86	0	40.15	382.63
	AN	40.62	100	84.03	0	19.33	12.73	0	11.29	10.73	0	40.3	387.6
	AN+FM	0	39.29	39.58	75	36.34	13.06	0	11.14	10.82	0	41.06	396.18
	TS	100	100	91.67	54.17	46.76	308.56	0	11.14	10.82	0	38.84	2320.86
Retrain	NA	0	100	92.36	20.83	0	79.72	0	98.36	88.51	35.12	0	1246.76
	TM	0	100	90.97	20.83	0	61.16	0	99.3	90.82	32.14	0	1515.63
	FM	0	100	95.14	12.5	0	55.28	0	99.23	91.35	39.96	0	2140.92
	AN	0	100	93.06	8.33	0	53.46	0	98.64	88.92	42.73	0	1185.3
	AN+FM	0	100	90.28	16.67	0	59.04	0	97.16	91.27	42.73	0	1952.47
	TS	0	100	94.44	16.67	0	1029.24	0	98.22	91.03	36.32	0	5078.88
SCRUB	NA	0	11.61	11.11	0	34.03	21.34	0	13.87	13.45	66.18	35.37	1107.27
	TM	0	11.61	11.11	100	53.01	21.37	0	16.83	16.97	50.87	30.86	1124.65
	FM	0	38.84	43.75	25	21.3	21.49	0	16.12	15.37	46.85	27.63	1145.55
	AN	0	28.57	29.86	58.33	37.73	21.47	0	15.42	14.79	72.96	34.79	1167.93
	AN+FM	0	34.82	36.81	37.5	24.77	21.3	0	12.26	14.87	71.39	35.02	1196.8
	TS	0	45.09	37.5	20.83	20.37	185.57	0.32	15.38	16.08	35.78	25.27	2386.63
SSD	NA	0	10.71	11.11	33.33	31.25	4.87	99.62	98.34	88.09	92.24	52.38	116.53
	TM	0	9.38	9.03	33.33	31.48	4.86	18.48	92.75	86.42	2.23	17.6	118.12
	FM	0	13.84	9.03	62.5	45.37	4.88	98.92	96.83	89.53	89.58	50.12	119.5
	AN	0	10.71	11.11	25	32.87	4.82	91.38	98.28	89.16	0	44.78	118.14
	AN+FM	0	12.95	11.11	62.5	41.67	5.01	66.38	94.86	90.6	0.11	36.55	118.1
	TS	0	10.27	12.5	41.67	35.65	22.81	0.75	92.76	86.85	2.93	12.77	250.74
UNSIR	NA	0	68.75	50	62.5	28.01	46.83	0.16	73.6	69.45	12.81	13.84	283.99
	TM	0	55.8	42.36	37.5	21.76	47.2	0.05	66.61	66.52	20.63	11.95	291.53
	FM	0	39.73	41.67	33.33	24.77	47.22	0	58.41	61.04	79.15	23.17	290.12
	AN	0	45.54	34.03	20.83	23.84	47.43	0.22	71.55	68.2	12.11	17.19	293.43
	AN+FM	0	41.07	34.03	41.67	27.08	47.65	0	13.25	15.95	54.67	29.09	304.68
	TS	0	64.73	50.69	20.83	15.97	201.53	0	34.62	34.61	18.57	24.72	1423.04

Table 5: Evaluation of Machine Unlearning Methods on the Us8K and GTZAN Datasets with the Audio Mamba Model

Datasets		**US8K**						**GTZAN**					
Unlearning Methods	**Aug**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**
Amnesiac	NA	0	99.9	83.78	2.75	11.06	350.96	0	100	74.65	21.79	1.98	186.11
	TM	0	96.74	78.32	1.75	9.24	355.86	0	100	78.12	33.33	8.16	186.41
	FM	0	95.44	83.64	7.75	11.67	359.49	0	99.18	73.26	53.85	14.57	186.16
	AN	0	99.7	82.85	2.5	8.94	354.28	0	100	72.22	25.64	8.77	186.65
	AN+FM	0	95.8	83.11	8.62	11.68	368.88	0	99.32	78.47	64.1	17.21	209.49
	TS	0	98.6	78.46	7.12	14.21	347.39	0	100	73.48	30.43	8.02	185.59
Bad Teacher	NA	0	93.61	83.78	0	11.98	378.27	0	97.15	73.61	0	7.02	199.67
	TM	0	88.66	75.4	0	10.8	387.67	0	95.65	78.82	0	4.11	199.85
	FM	0	88.74	83.11	0	14.43	384.56	0	93.75	76.74	0	4.54	199.62
	AN	0	91.53	77.93	0	11.41	371.62	0	92.26	65.97	0	4.87	201.59
	AN+FM	1.38	90.7	84.57	0	14.52	379.69	0	90.9	70.83	0	6.94	202.17
	TS	0	95	77.79	0	16.81	397.73	0	76.33	67.23	0	7.68	196.19
Boundary	NA	0	16.6	19.41	40.38	22.94	93.27	34.64	32.34	31.6	25.64	29.16	42.5
	TM	67.88	1.15	3.06	65	62.12	95.79	45.71	80.43	59.38	64.1	38.97	42
	FM	0	15.27	18.35	67	31.44	96.89	2.5	63.72	61.46	66.67	23.61	40.29
	AN	0	12.5	12.23	29.12	24.18	97.73	45	68.34	60.07	53.85	37.22	40.38
	AN+FM	0	12.17	11.04	85.62	39.7	95.84	12.68	63.32	57.64	66.67	29.24	40.36
	TS	0	13.43	12.9	59.38	26.65	95.09	44.25	56.03	55.4	63.77	39.91	37.98
Fisher	NA	1	87.42	78.06	10.12	10.22	5746.21	1.25	81.11	62.15	6.41	9.12	629.27
	TM	4	80.03	73.94	12	8.62	5812.97	1.43	74.86	62.5	11.54	5.65	633.82
	FM	9.62	85.56	82.31	10	14.57	5874.87	1.25	86.55	75	1.28	3.95	633.92
	AN	0.62	87.25	77.66	6.75	9.46	5903.9	2.5	78.8	57.64	11.54	9.76	646
	AN+FM	2.25	84.61	79.65	3.75	15.2	5921.27	7.68	80.57	63.19	0	12.04	642.58
	TS	0.88	88.72	72.07	3.75	17.76	5790.35	7.75	63.72	57.95	10.14	9.98	710.38
Gradient Ascent	NA	0	13.03	12.37	0	35.16	147.21	97.14	97.69	72.57	80.77	54.71	70.71
	TM	0	13.09	12.77	0	31.68	147.81	2.5	57.88	49.31	28.21	15.96	70.7
	FM	0	13.09	12.77	0	37.88	147.97	87.14	95.11	78.82	41.03	40.27	70.69
	AN	0	13.09	12.77	0	33.13	147.78	44.29	95.24	74.65	2.56	15.88	74.41
	AN+FM	0	13.09	12.77	0	38	152.18	40.54	84.24	68.06	53.85	30.78	75.09
	TS	0	12.17	11.04	0	39.06	143.32	36.75	73.67	61.08	28.99	23.92	69.29
Retrain	NA	0	99.25	82.85	35	0	384.74	0	100	76.74	17.95	0	237.71
	TM	0	98.7	82.18	25.62	0	384.48	0	100	76.74	10.26	0	281.26
	FM	0	97.61	86.04	40.38	0	436.34	0	98.78	74.65	11.54	0	388.31
	AN	0	99.9	83.91	28.25	0	433.79	0	100	76.74	3.85	0	290.91
	AN+FM	0	97.68	85.64	41.12	0	434.43	0	99.46	78.82	12.82	0	425.45
	TS	0	99.14	80.85	47.38	0	431.62	0	100	78.69	11.59	0	235.49
SCRUB	NA	0	15.43	15.29	10	30.85	418.35	0	65.08	49.31	14.1	10.43	204.26
	TM	0	17.43	20.21	100	45.45	421.34	0	85.46	69.1	8.97	2.97	204.81
	FM	0	19.45	24.07	76.88	32.82	424.21	0	44.7	39.24	14.1	12.66	204.66
	AN	0	28.97	27.79	70.75	32.87	418.8	0	48.1	36.46	39.74	25.39	204.73
	AN+FM	0	29.75	31.12	84.75	32.72	425.82	0	65.22	55.56	33.33	14.59	204.36
	TS	0	32.11	30.59	28.88	22.92	421.94	0	53.4	49.72	15.94	11.11	206.12
SSD	NA	0	15.64	17.55	37.5	22.6	43.54	0	42.93	30.9	100	42.63	21.32
	TM	0	5.95	5.32	32.5	27.91	45.96	0	12.5	13.54	57.69	36.88	21.36
	FM	1.88	44.74	48.67	14	21.87	46.04	6.43	56.66	44.44	5.13	14.35	21.6
	AN	0	30.18	31.38	62.62	28.97	44.88	0	9.65	13.54	47.44	35.59	20.34
	AN+FM	0	13.13	13.03	42.38	24.62	45.34	29.29	18.48	11.81	67.95	50.48	21.82
	TS	0	12.8	15.16	62.25	26.86	45.84	28.75	30.05	28.5	43.48	36.94	21.84
UNSIR	NA	0	73.59	61.84	40.62	8.88	186.91	0	57.2	50	55.13	21.31	6192.96
	TM	0	73.69	60.24	42.75	13.02	190.45	0	24.46	22.22	33.33	25.86	6212.88
	FM	0	73.83	67.42	28.62	10.12	195.09	0	60.05	54.86	42.31	16.85	6247.07
	AN	0	78.46	63.3	30.5	7.62	196.92	0	49.32	41.67	46.15	25.79	6262.81
	AN+FM	0	52.65	52.93	46.38	12.65	198.93	0	19.29	15.62	25.64	25.34	6331
	TS	0	75.87	59.97	25.88	14.13	191.57	0	49.4	40.81	37.68	21.32	8257.84

6. Mask-based augmentations preserve UNA for many unlearning methods. TS, AN, FM, and NA frequently appear in the Best UNA category, indicating these perturbations help maintain high accuracy on non-forgotten classes even after model unlearning.
7. TM (Time Mask) is repeatedly associated with weak forgetting and privacy performance. TM often appears as Worst AGP or Worst MIA (e.g., Fisher for GTZAN, BadTeacher for ESC10, UNSIR for ESC10), showing that masking temporal segments harms unlearning fidelity by removing essential temporal cues needed by ResNet-18.
8. UNSIR exhibits strong and stable performance across different augmentations and datasets. UNSIR frequently reports strong AGP/UNA performance (e.g., ESC10–TS, GTZAN–TS, GSCD–FM), suggesting that its structured regularization makes it less vulnerable to augmentation-induced instability.
9. The TS, AN, and FM (individually) generally contribute to better AGP, MIA, and UNA, whereas AN+FM and TM usually degrade performance. This indicates that moderate perturbations benefit the convolutional architecture, while combined or severe temporal masking disrupts feature extraction crucial for stable unlearning.

5.4. ***Ablation Study***

5.4.1. ***Entropy Analysis***

To better understand the mechanism by which audio augmentation influences unlearning, we analysed the Shannon entropy of the model's softmax output distributions. Entropy serves as a proxy for model uncertainty. In an ideal unlearning scenario, the model should exhibit high entropy (high uncertainty) on the Forget Set (D_f) indicating effective removal of information while maintaining low or stable entropy (high confidence) on the Retain Set (D_r). We define the entropy *H(P)* for a probability distribution $\mathbf{P} = \{p_1, \ldots, p_C\}$ over C classes as:

$$H(\mathbf{P}) = -\sum_{i=1}^{C} p_i \log(p_i) \tag{1}$$

We compared the mean entropy of models unlearned with and without augmentation across four distinct configurations (Figure 3): AST (GTZAN), AST (GSCD), Audio Mamba (US8K), and ResNet-18 (GSCD).

Table 6: Evaluation of Machine Unlearning Methods on the ESC-10 and GSCD Datasets with the ResNet-18 Model

Datasets		**ESC10**						**GSCD**					
Unlearning Methods	**Aug**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**
Amnesiac	NA	0	100	88.89	4.17	1.39	6.76	0	99.95	90.34	54.13	10.08	211.2
	TM	0	100	90.97	45.83	12.96	10.65	0	99.27	91.74	47.67	5.52	217.96
	FM	0	98.66	89.58	16.67	4.4	7.44	0	99.31	92.69	51.95	8.42	207.77
	AN	0	100	84.03	83.33	27.78	12.98	0	99.84	92	55.86	10.53	199.78
	AN+FM	0	98.21	88.89	45.83	15.74	7.78	0	95.37	92.37	55.37	6.06	198.1
	TS	0	100	88.19	41.67	11.34	214.43	0	99.44	95.04	15.64	10.65	550.95
Bad Teacher	NA	31.25	100	88.89	0	10.42	6.19	0.16	97.71	90.62	0	8.11	205.88
	TM	18.75	100	92.36	0	9.95	7.06	0	96.93	90.84	0	11.39	217.59
	FM	28.12	100	87.5	0	11.23	6.61	0.32	97.95	92.3	0	9.13	217.51
	AN	3.12	99.55	84.72	0	1.27	6.34	0.05	97.44	90.79	0	8.9	202.61
	AN+FM	21.88	98.66	89.58	0	7.52	7.69	0.32	94.03	93.02	0	13.01	209.9
	TS	9.38	99.55	85.42	0	7.06	144.23	0	98.17	94.96	0	15.83	1793.71
Boundary	NA	34.38	91.07	73.61	0	16.55	1.6	66.54	97.4	86.95	21.23	24.31	37.08
	TM	28.12	88.39	72.92	0	17.71	1.7	54.96	92.84	87.23	25.73	22.33	38.78
	FM	28.12	85.27	79.86	4.17	12.38	1.69	47.79	96.39	91.55	14.28	20.44	37.95
	AN	25	63.84	73.61	0	11.81	1.71	50.59	95	88.64	17.75	20.55	34.84
	AN+FM	31.25	82.59	75	4.17	16.9	2.37	43.75	85.78	89.22	79.75	29.82	34.44
	TS	15.62	84.38	71.53	4.17	12.38	17.64	36.37	96.49	92.78	8.85	25.56	191.63
Fisher	NA	6.25	87.05	71.53	0	7.87	17.92	13.63	95.59	86.33	2.06	13.27	1799.04
	TM	9.38	91.52	84.03	0	7.75	18.36	5.33	94.89	88.77	5.27	12.1	1888.8
	FM	3.12	95.54	82.64	0	4.51	18.33	15.84	96.08	89.4	1.74	14.69	1869.33
	AN	37.5	96.88	74.31	0	15.74	20.01	4.26	95.18	87.09	4.23	10.13	1830.68
	AN+FM	3.12	95.09	79.86	4.17	5.9	19.14	9	92.25	90.99	2.17	15.85	1845.97
	TS	0	95.54	81.25	0	5.32	42.78	4.09	97.11	93.62	2.55	16.62	2144.44
Gradient Ascent	NA	28.12	89.73	72.92	0	14.7	3.49	0	86.19	79.15	48.48	11.93	83.04
	TM	28.12	88.39	72.22	0	17.94	4.41	0	88.22	82.34	59.55	12.61	96.02
	FM	21.88	86.61	81.25	0	11.23	3.96	0	85.31	79.89	63.08	16.2	99.19
	AN	21.88	61.16	74.31	4.17	11.92	4.28	0	79.13	73.28	59.72	18.05	91.7
	AN+FM	21.88	83.93	70.14	4.17	15.39	4.38	0	66	73.17	36.86	7.23	103.61
	TS	12.5	82.14	72.92	0	12.27	281.16	0	91.07	87.83	62.7	7.45	2199.41
Retrain	NA	0	100	88.89	0	0	29.54	0	99.86	90.51	24.05	0	278.99
	TM	0	100	89.58	8.33	0	28.53	0	99.62	92.82	32.19	0	439.24
	FM	0	99.11	88.89	4.17	0	30.91	0	99.34	92.39	26.98	0	497.79
	AN	0	100	84.03	0	0	37.91	0	99.86	92.59	24.86	0	521.26
	AN+FM	0	98.66	90.28	0	0	34.28	0	98.1	93.45	38.27	0	760.89
	TS	0	100	88.89	8.33	0	1523.58	0	99.47	94.7	47.23	0	5572.67
SCRUB	NA	0	100	88.89	12.5	4.17	6.08	0	13.65	13.3	81.54	44.9	197.87
	TM	0	100	89.58	20.83	4.17	3.86	0	12.83	11.47	0.92	37.54	203.94
	FM	0	92.41	86.81	54.17	17.36	6.49	0	16.75	17.75	34.58	27.41	194.12
	AN	0	100	84.03	29.17	9.72	4.95	0	13.89	14.43	81.92	45.07	206.82
	AN+FM	0	95.09	88.89	58.33	19.91	6.13	0	12.26	11.96	33.39	28.79	196.41
	TS	0	99.11	84.72	29.17	8.33	168.15	0	11.7	11.49	20.74	36.57	1613.88
SSD	NA	0	2.23	4.17	20.83	35.19	0.81	93.91	99.42	90.26	17.86	33.45	18.03
	TM	0	100	92.36	0	3.7	0.9	40.09	98.53	91.72	1.3	24.02	17.51
	FM	0	91.52	75.69	0	5.79	0.8	99.68	98.8	92.5	96.63	56.48	17.3
	AN	0	10.71	16.67	62.5	43.29	0.89	0	98.35	91.35	7.82	6.09	19.28
	AN+FM	0	97.77	92.36	0	0.69	0.96	98.76	95.26	93.19	11.13	42.05	19.84
	TS	0	9.82	11.11	54.17	41.2	17.97	98.38	99.02	95	66.83	39.43	203.35
UNSIR	NA	3.12	99.11	88.89	45.83	16.32	15.92	0	94.91	92.41	32.25	3.36	84.06
	TM	0	98.21	89.58	62.5	18.06	15.04	0.16	94.36	93.53	39.58	2.75	93.7
	FM	3.12	91.52	73.61	41.67	18.63	17.23	0.05	93.49	93.16	29.15	1	89.85
	AN	15.62	99.55	65.28	45.83	26.74	15.15	0	94.53	93.25	42.62	6.14	99.53
	AN+FM	3.12	83.04	78.47	45.83	20.25	18.46	0.54	87.74	93.08	43.27	1.97	102.6
	TS	3.12	98.66	86.81	33.33	10.07	163.86	0	96.2	95	29.86	5.89	1434.23

Table 7: Evaluation of Machine Unlearning Methods on the Us8K and GTZAN Datasets with the ResNet-18 Model

Datasets		**US8K**						**GTZAN**					
Unlearning Methods	**Aug**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**	**UNA**	**RMA**	**TEA**	**MIA**	**AGP**	**RTE**
Amnesiac	NA	0	99.92	75.93	1.88	7.34	77.1	0	100	80.9	67.95	22.65	79.05
	TM	0	98.73	75.27	3.5	7.15	98.02	0	100	75.35	69.23	24.07	65.67
	FM	0	90.81	76.86	1.38	7.92	89.97	0	98.78	78.47	97.44	30.68	98.14
	AN	0	99.94	76.2	1.75	7.47	83.29	0	100	80.9	58.97	20.58	59.44
	AN+FM	0	92.64	78.59	0.88	6.34	88.67	0	99.18	78.47	42.31	9.86	53.01
	TS	0	99.98	70.88	2.5	4.87	74.5	0	100	76.61	47.83	16.18	66.1
Bad Teacher	NA	0.62	91.85	71.94	0	7.82	73.03	3.75	100	75.35	0	3.1	56.49
	TM	1.5	93.35	73.94	0	9.26	73.73	9.11	100	78.82	0	4.58	56.7
	FM	0.62	82.61	71.41	0	10.41	75.8	5.18	95.79	78.47	0	6.77	56.63
	AN	1.25	94.13	74.07	0	8.91	69.98	18.04	99.86	76.39	0	6.59	58.79
	AN+FM	1.62	87.32	75	0	8.19	75.48	7.86	97.69	76.04	0	8.59	59.43
	TS	0.12	93.45	72.07	0	5.34	67.27	13.75	100	75.57	0	7.1	56.14
Boundary	NA	52.12	80.17	63.43	25.88	22.2	12.78	6.61	58.97	46.88	48.72	29.78	12.06
	TM	24.75	57.13	56.12	31.38	17.49	13.95	2.5	66.71	50.35	58.97	29.82	12.05
	FM	34.5	50.51	60.51	75.12	33.79	13.94	1.25	46.6	44.44	37.18	22.35	12.05
	AN	63.12	85.62	67.95	28.38	25.2	13.74	9.11	66.98	48.26	43.59	27.52	12.04
	AN+FM	52.88	63.84	64.49	66.38	37.82	13.88	0	33.7	36.81	10.26	15.63	12.04
	TS	30.75	80.85	56.91	14.12	15.89	13.66	3.75	89.92	65.72	4.35	6.56	12.48
Fisher	NA	6	87.94	70.61	1.88	9.43	698.98	1.25	95.38	77.78	0	1.46	66.03
	TM	5.12	89.37	72.21	7	8.71	680.8	1.25	95.24	78.82	0	1.97	68.43
	FM	1.12	79.61	75	6.75	7.13	684.17	12.68	94.84	77.43	0	9.61	70.41
	AN	6.75	91.63	73.27	3.75	9.76	703.98	2.5	96.2	78.82	1.28	1.49	67.39
	AN+FM	1.12	86	73.54	5.25	6.76	690.73	11.61	92.93	81.6	0	9.15	69.9
	TS	0.75	88.68	67.69	2.12	6.3	700.3	0	92.23	70.36	1.45	3.77	67.51
Gradient Ascent	NA	0	85.63	70.61	12.38	3.93	27.65	3.75	86.28	66.67	1.28	6.42	20.88
	TM	0	69.57	64.76	53.38	13.69	31.03	7.5	96.88	69.79	0	7.06	21.34
	FM	0	58.4	65.03	59.62	15.62	33.24	7.5	82.47	68.4	3.85	9.61	21.37
	AN	0	88.55	74.2	12.62	4.24	29.86	6.25	94.29	70.83	1.28	4.94	25.34
	AN+FM	0	72.03	69.28	76	21.81	35.62	3.75	89.95	71.88	6.41	6.48	26.63
	TS	0	89.17	66.49	11.62	3.29	25.73	7.75	96.33	74.53	1.45	4.96	19.04
Retrain	NA	0	99.92	73.4	21.38	0	121.79	0	100	80.9	0	0	186.22
	TM	0	99.08	76.33	23.88	0	134.56	0	100	80.9	2.56	0	190.92
	FM	0	93.83	77.13	24.88	0	169.33	0	99.59	83.33	10.26	0	197.44
	AN	0	99.92	75.8	23.75	0	126.63	0	100	78.12	0	0	258.28
	AN+FM	0	97.73	78.32	19.62	0	308.75	0	99.59	79.86	14.1	0	264.56
	TS	0	100	72.47	15.5	0	115.7	0	100	80.21	2.9	0	104.76
SCRUB	NA	0	18.01	20.61	1.25	24.31	66.77	0	99.05	82.64	20.51	7.42	55.44
	TM	0	23.59	27.53	34	19.64	68.53	0	98.51	77.43	37.18	12.7	55.43
	FM	0	27.94	29.26	98.25	40.42	72.58	0	91.3	83.33	24.36	4.7	55.39
	AN	0	40.91	31.38	25.38	15.35	66.86	0	99.18	81.6	20.51	8	55.46
	AN+FM	0	22.63	21.68	26.75	21.26	70.19	0	93.89	76.74	16.67	1.9	55.51
	TS	0	20.6	12.9	41.12	28.4	67.1	100	100	81.25	14.49	37.55	56.56
SSD	NA	6.75	93.91	73.14	0.12	9.42	5.87	0	34.92	28.12	12.82	21.87	5.27
	TM	18.5	94.13	76.73	0.88	13.97	6.5	0	100	77.43	0	2.01	4.59
	FM	0	33.29	32.45	16.38	17.73	7.5	98.57	98.37	74.65	41.03	46.01	4.67
	AN	0	95.87	75	3.12	7.14	6.2	0	23.51	22.57	88.46	48.01	4.6
	AN+FM	87.62	88.41	73.27	41	38.02	6.83	15	99.32	79.17	0	9.93	6.19
	TS	0	95.9	73.8	1.38	5.15	6.09	0	10.33	18.56	57.97	38.91	4.77
UNSIR	NA	0	89.03	73.94	40.38	6.51	45.13	0	98.91	81.94	15.38	5.48	168.35
	TM	0	88.25	74.07	40.5	6.3	48.9	0	99.46	82.99	23.08	7.53	170.65
	FM	0	85.31	78.72	45	7.24	53.36	9.29	85.87	71.53	16.67	9.17	151.4
	AN	0	89.19	74.73	28.12	1.81	47.7	0	97.69	77.78	14.1	4.82	156.01
	AN+FM	0	85.38	76.6	44.5	8.87	51.7	0	84.78	81.94	41.03	9.67	157.28
	TS	0.25	88.34	68.62	35.75	8.12	43.12	0	99.18	78.69	10.14	2.92	172.46

Table 8: Comprehensive Summary of the Performance of Different Audio Augmentation Methods During Unlearning with the AST Model

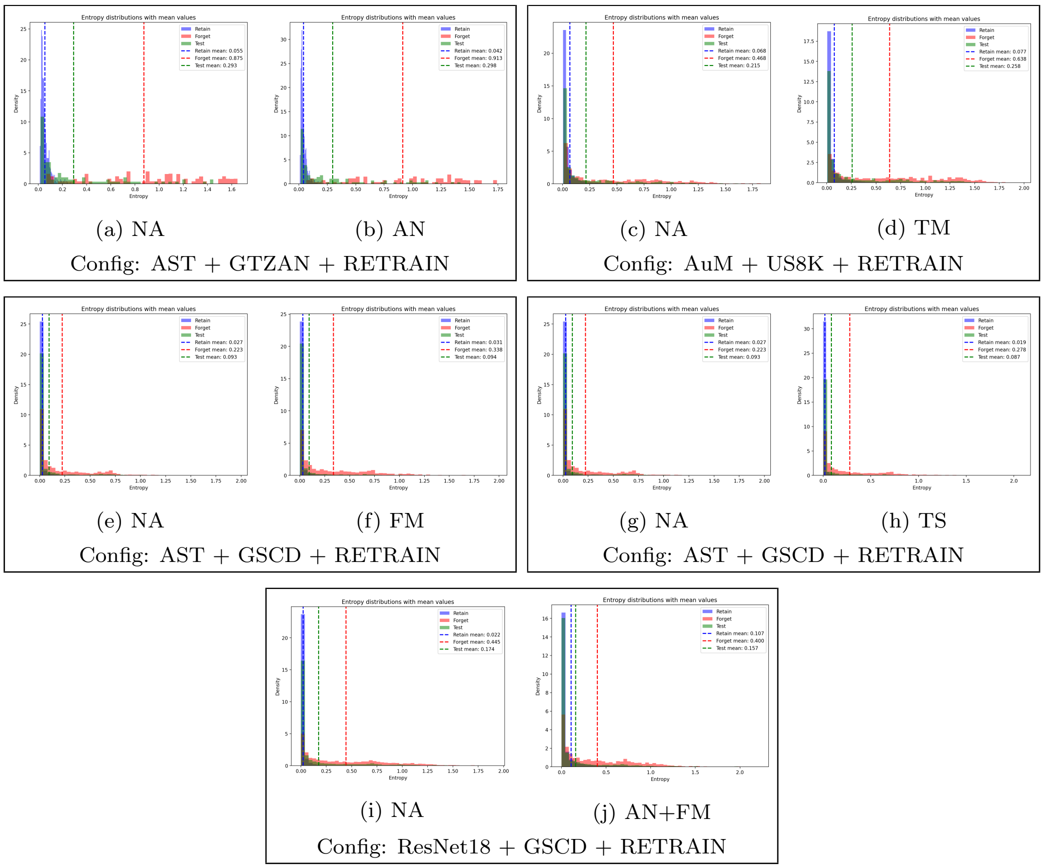
Dataset	Unlearning Method	Best AGP	Worst AGP	Best MIA	Worst MIA	Best UNA	Worst UNA
ESC10	Amnesiac	NA	AN+FM	TS	AN+FM, FM	ALL	-
	BadTeacher	FM	TS	ALL	-	AN, FM, NA, TM	TS
	Boundary	TS	NA	AN	TM	TS	NA
	Fisher	AN, AN+FM	TM	ALL	-	AN, AN+FM, TS	TM
	Gradient Ascent	AN+FM	NA	FM	NA	AN, AN+FM, TM, TS	FM, NA
	Retrain	ALL	-	AN, AN+FM	TM	ALL	-
	SCRUB	TS	NA	AN+FM, TS	NA	ALL	-
	SSD	FM	AN+FM	AN, FM	AN+FM	ALL	-
	UNSIR	TM	NA	AN	FM	ALL	-
GTZAN	Amnesiac	AN	TM	AN	TM	ALL	-
	BadTeacher	TM	TS	ALL	-	AN, FM, NA, TM, TS	AN+FM
	Boundary	FM	TS	FM, TM	TS	NA	AN+FM
	Fisher	AN+FM	TS	AN+FM, TS	NA, TM	AN+FM	TS
	Gradient Ascent	FM	AN	FM	AN+FM	ALL	-
	Retrain	ALL	-	AN	FM	ALL	-
	SCRUB	TS	AN	TS	AN	TS	AN, FM
	SSD	NA	AN+FM	FM, NA, TS	AN+FM	AN, FM, NA, TM, TS	AN+FM
	UNSIR	TS	TM	TS	NA	ALL	-
GSCD	Amnesiac	FM	TM	TS	AN+FM	ALL	-
	BadTeacher	FM	TM	ALL	-	TM	TS
	Boundary	TS	FM	AN+FM	FM	TS	NA
	Fisher	FM	TM	TS	TM	AN+FM, FM	NA
	Gradient Ascent	TM	FM	TS	TM	TM	AN+FM
	Retrain	ALL	-	FM	TM	ALL	-
	SCRUB	AN	FM	AN+FM	TM	AN	TM
	SSD	NA	AN+FM	AN	AN+FM	NA, TM	AN+FM
	UNSIR	AN	AN+FM	TS	AN+FM	ALL	-
US8K	Amnesiac	NA	AN	NA	AN+FM	ALL	-
	BadTeacher	NA	AN	ALL	-	ALL	-
	Boundary	NA	TS	FM	AN	TM	TS
	Fisher	NA	TM	AN, AN+FM	FM	AN+FM	TM
	Gradient Ascent	TM	AN+FM	TM	NA	AN, FM, NA, TM, TS	AN+FM
	Retrain	ALL	-	NA	FM	ALL	-
	SCRUB	FM	NA	TS	AN+FM	TS	AN
	SSD	TM	AN+FM	NA	AN+FM	NA	AN+FM
	UNSIR	TS	NA	TS	TM	ALL	-

Table 9: Comprehensive Summary of the Performance of Different Audio Augmentation Methods During Unlearning with the Audio Mamba Model

Dataset	Unlearning Method	Best AGP	Worst AGP	Best MIA	Worst MIA	Best UNA	Worst UNA
ESC10	Amnesiac	NA	AN+FM	NA	AN+FM	ALL	-
	BadTeacher	AN	NA	ALL	-	AN, NA, TM, TS	AN+FM, FM
	Boundary	FM	TM	AN+FM, FM, NA	AN, TM	NA	AN
	Fisher	AN+FM	NA	AN+FM	NA	TS	NA
	Gradient Ascent	AN	TS	AN	AN+FM	AN+FM	NA, TM, TS
	Retrain	ALL	-	AN	NA, TM	ALL	-
	SCRUB	TS	TM	NA	TM	ALL	-
	SSD	NA	FM	AN	AN+FM, FM	ALL	-
	UNSIR	TS	NA	AN, TS	NA	ALL	-
GTZAN	Amnesiac	NA	AN+FM	NA	AN+FM	ALL	-
	BadTeacher	TM	TS	ALL	-	ALL	-
	Boundary	FM	TS	NA	AN+FM, FM	FM	TM
	Fisher	FM	AN+FM	AN+FM	AN, TM	FM, NA	TS
	Gradient Ascent	AN	NA	AN	NA	TM	NA
	Retrain	ALL	-	AN	NA	ALL	-
	SCRUB	TM	AN	TM	AN	ALL	-
	SSD	FM	AN+FM	FM	NA	AN, NA, TM	AN+FM
	UNSIR	FM	TM	AN+FM	NA	ALL	-
GSCD	Amnesiac	TM	AN+FM	NA	TM	ALL	-
	BadTeacher	TM	AN+FM	ALL	-	AN, AN+FM, NA, TM	FM, TS
	Boundary	TS	NA	AN	NA	TS	NA
	Fisher	TM	TS	NA	AN+FM	AN+FM, TM	TS
	Gradient Ascent	TM	AN+FM	ALL	-	ALL	-
	Retrain	ALL	-	TM	AN, AN+FM	ALL	-
	SCRUB	TS	NA	TS	AN	AN, AN+FM, FM, NA, TM	TS
	SSD	TS	NA	AN	NA	TS	NA
	UNSIR	TM	AN+FM	AN	FM	AN+FM, FM, TS	AN
Us8K	Amnesiac	AN	TS	TM	AN+FM	ALL	-
	BadTeacher	TM	TS	ALL	-	AN, FM, NA, TM, TS	AN+FM
	Boundary	NA	TM	AN	AN+FM	AN, AN+FM, FM, NA, TS	TM
	Fisher	TM	TS	AN+FM, TS	TM	AN	FM
	Gradient Ascent	TM	TS	ALL	-	ALL	-
	Retrain	ALL	-	TM	TS	ALL	-
	SCRUB	TS	TM	NA	TM	ALL	-
	SSD	FM	AN	FM	AN	AN, AN+FM, NA, TM, TS	FM
	UNSIR	AN	TS	TS	AN+FM	ALL	-

Table 10: Comprehensive Summary of the Performance of Different Audio Augmentation Methods During Unlearning with the ResNet-18 Model

Dataset	Unlearning Method	Best AGP	Worst AGP	Best MIA	Worst MIA	Best UNA	Worst UNA
ESC10	Amnesiac	NA	AN	NA	AN	ALL	-
	BadTeacher	AN	FM	ALL	-	AN	NA
	Boundary	AN	TM	AN, NA, TM	AN+FM, FM, TS	TS	NA
	Fisher	FM	AN	AN, FM, NA, TM, TS	AN+FM	TS	AN
	Gradient Ascent	FM	TM	FM, NA, TM, TS	AN, AN+FM	TS	NA, TM
	Retrain	ALL	-	AN, AN+FM, NA	TM, TS	ALL	-
	SCRUB	NA, TM	AN+FM	NA	AN+FM	ALL	-
	SSD	AN+FM	AN	AN+FM, FM, TM	AN	ALL	-
	UNSIR	TS	AN	TS	TM	TM	AN
GTZAN	Amnesiac	AN+FM	FM	AN+FM	FM	ALL	-
	BadTeacher	NA	AN+FM	ALL	-	NA	AN
	Boundary	TS	TM	TS	TM	AN+FM	AN
	Fisher	NA	FM	AN+FM, FM, NA, TM	TS	TS	FM
	Gradient Ascent	AN	FM	TM	AN+FM	AN+FM, NA	TS
	Retrain	ALL	-	AN, NA	AN+FM	ALL	-
	SCRUB	AN+FM	TS	TS	TM	AN, AN+FM, FM, NA, TM	TS
	SSD	TM	AN	AN+FM, TM	AN	AN, NA, TM, TS	FM
	UNSIR	TS	AN+FM	TS	AN+FM	AN, AN+FM, NA, TM, TS	FM
GSCD	Amnesiac	TM	TS	TS	AN	ALL	-
	BadTeacher	NA	TS	ALL	-	TM, TS	AN+FM, FM
	Boundary	FM	AN+FM	TS	AN+FM	TS	NA
	Fisher	AN	TS	FM	TM	TS	FM
	Gradient Ascent	AN+FM	AN	AN+FM	FM	ALL	-
	Retrain	ALL	-	NA	TS	ALL	-
	SCRUB	FM	AN	TM	AN	ALL	-
	SSD	AN	FM	TM	FM	AN	FM
	UNSIR	FM	AN	FM	AN+FM	AN, NA, TS	AN+FM
US8K	Amnesiac	TS	FM	AN+FM	TM	ALL	-
	BadTeacher	TS	FM	ALL	-	TS	AN+FM
	Boundary	TS	AN+FM	TS	FM	TM	AN
	Fisher	TS	AN	NA	TM	TS	AN
	Gradient Ascent	TS	AN+FM	TS	AN+FM	ALL	-
	Retrain	ALL	-	TS	FM	ALL	-
	SCRUB	AN	FM	NA	FM	ALL	-
	SSD	TS	AN+FM	NA	AN+FM	AN, FM, TS	AN+FM
	UNSIR	AN	AN+FM	AN	FM	AN, AN+FM, FM, NA, TM	TS



(a) NA (b) AN
Config: AST + GTZAN + RETRAIN

(c) NA (d) TM
Config: AuM + US8K + RETRAIN

(e) NA (f) FM
Config: AST + GSCD + RETRAIN

(g) NA (h) TS
Config: AST + GSCD + RETRAIN

(i) NA (j) AN+FM
Config: ResNet18 + GSCD + RETRAIN

Figure 3: Entropy Analysis of Forget and Retain Set

1. **Augmentation Increases Unlearning Uncertainty:** A key finding from our ablation study is that effective audio augmentation consistently increases the entropy of the D_F, indicating stronger disruption of the targeted representations. For instance, when unlearning is performed on Audio-Mamba with Time Masking (TM) on the US8K dataset, the entropy of the D_F rises sharply compared to the no-augmentation scenario (see Figure 3). A similar trend appears in the AST model, where Time Shifting (TS) augmentation during unlearning on the GSCD dataset produces a marked entropy increase. This systematic rise in entropy explains the observed improvements across multiple unlearning metrics when suitable audio augmentations are applied.

2. **Augmentation Preserves Generalization:**Another important observation from our ablation study is that, in successful unlearning configurations, the entropy of the D_R remains stable or even slightly decreases. This behaviour indicates that augmentation-guided unlearning helps transition the model from memorisation toward generalisation. The model preserves high confidence on the data it is meant to retain, demonstrating that the unlearning process is selective and does not trigger catastrophic forgetting.
3. **Entropy Explains Failure Cases:** Our analysis also explains why the AN+FM augmentation often failed, especially for ResNet-18. As shown in Figure 3, this combination caused the opposite behaviour of what good unlearning should achieve: the D_F showed higher entropy, while the D_R showed lower entropy. This means the model became confused rather than selectively forgetting. Instead of removing information, the model seemed to latch onto noisy patterns in the forget set and lost stability on the retain set. This confusion directly matches the poor RMA scores reported in Table 6.

5.5. ***Discussion***

Table 8, Table 9, and Table 10 present a comprehensive summary of the extensive experiments performed across four audio classification datasets. These tables offer a comparative view of how various data augmentation techniques influence the unlearning performance of the AST, Audio Mamba, and ResNet-18 models. Across all three architectures AST, ResNet-18, and Audio-Mamba, it is observed that incorporating carefully selected audio data augmentation substantially enhances the performance of machine unlearning methods, helping to narrow the gap with the fully retrained baseline. While Retrain remains the most stable and highest-performing approach across models and datasets, augmentation-guided unlearning consistently improves AGP, UNA, and MIA scores for most methods. Techniques such as Time Masking (TM), Additive Noise (AN), and Frequency Masking (FM) often yield significant gains by regularizing the model's temporal and spectral representations, whereas aggressive combinations like AN+FM can destabilize unlearning, particularly in AST and Audio-Mamba. ResNet-18 benefits most consistently from augmentations due to its convolutional inductive biases, while AST and Audio-Mamba exhibit dataset-dependent sensitivity but still achieve notable improvements when augmentations are aligned with their architectural characteristics. Overall, the analysis demonstrates that augmentation-driven unlearning not only enhances the effectiveness of forgetting and privacy robustness but also reduces the discrepancy between unlearned and fully retrained models, highlighting data augmentation as a key factor for reliable, high-fidelity machine unlearning in audio classification systems.

6. Conclusion

In this work, we conducted the first comprehensive empirical investigation into the role of audio data augmentation in machine unlearning for audio models. By evaluating nine state-of-the-art unlearning methods across three widely used architectures, Viz. AST, Audio-Mamba, and ResNet-18—and four benchmark datasets (Google Speech Commands V1, ESC-10, GTZAN, and US8K), we demonstrated that incorporating well-chosen audio augmentation strategies substantially improves the effectiveness of machine unlearning. Our results show that augmentations such as time masking and additive noise consistently reduce the performance gap between retrained and unlearned models, enhance class-wise forgetting behavior, and mitigate the memorization effects that typically hinder audio unlearning. This study establishes audio data augmentation as a crucial component for building reliable, privacy-preserving, and high-fidelity audio unlearning pipelines. The natural extension of our work can be listed as follows:

- The first future work will be to extend augmentation-guided unlearning to broader audio tasks such as speech recognition, speaker identification, music genre classification, and multilingual speech scenarios.
- Another research direction involves designing augmentation-aware unlearning techniques for self-supervised audio models, large-scale audio transformers, and real-time or streaming audio systems.
- Developing theoretical foundations that explain the role of audio augmentation in stabilizing and improving unlearning remains an important challenge for advancing robust and generalizable audio unlearning frameworks.

References

[1] A. Mantelero, The eu proposal for a general data protection regulation and the roots of the 'right to be forgotten', Computer Law & Security Review 29 (3) (2013) 229–235.

[2] N. Li, C. Zhou, Y. Gao, H. Chen, Z. Zhang, B. Kuang, A. Fu, Machine unlearning: Taxonomy, metrics, applications, challenges, and prospects, IEEE Transactions on Neural Networks and Learning Systems (2025).

[3] M. H. Huang, L. G. Foo, J. Liu, Learning to unlearn for robust machine unlearning, in: European Conference on Computer Vision, Springer, 2024, pp. 202–219.

[4] W. Wang, Z. Tian, C. Zhang, S. Yu, Machine unlearning: A comprehensive survey, arXiv preprint arXiv:2405.07406 (2024).

[5] A. Golatkar, A. Achille, S. Soatto, Eternal sunshine of the spotless net: Selective forgetting in deep networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 9304–9312.

[6] V. Suriyakumar, A. C. Wilson, Algorithms that approximate data removal: New results and limitations, Advances in Neural Information Processing Systems 35 (2022) 18892–18903.

[7] P. Maini, Z. Feng, A. Schwarzschild, Z. C. Lipton, J. Z. Kolter, Tofu: A task of fictitious unlearning for llms, arXiv preprint arXiv:2401.06121 (2024).

[8] J. Ji, Y. Liu, Y. Zhang, G. Liu, R. R. Kompella, S. Liu, S. Chang, Reversing the forget-retain objectives: An efficient llm unlearning framework from logit difference, Advances in Neural Information Processing Systems 37 (2024) 12581–12611.

[9] A. M. Kassem, O. A. M. Mahmoud, S. Saad, Preserving privacy through dememorization: An unlearning technique for mitigating memorization risks in language models (2023).

[10] A. Koudounas, C. Savelli, F. Giobergia, E. Baralis, "Alexa, can you forget me?" Machine Unlearning Benchmark in Spoken Language Understanding, in: Interspeech 2025, 2025, pp. 1768–1772. doi:10.21437/Interspeech.2025-2607.

[11] J. Cheng, H. Amiri, Speech Unlearning, in: Interspeech 2025, 2025, pp. 3209–3213. doi:10.21437/Interspeech.2025-2412.

[12] Z. Di, Z. Zhu, J. Jia, J. Liu, Z. Takhirov, B. Jiang, Y. Yao, S. Liu, Y. Liu, Label smoothing improves machine unlearning, arXiv preprint arXiv:2406.07698 (2024).

[13] J. Jia, J. Liu, P. Ram, Y. Yao, G. Liu, Y. Liu, P. Sharma, S. Liu, Model sparsity can simplify machine unlearning, Advances in Neural Information Processing Systems 36 (2023) 51584–51605.

[14] K. Zhao, M. Kurmanji, G.-O. Bărbulescu, E. Triantafillou, P. Triantafillou, What makes unlearning hard and what to do about it, Advances in Neural Information Processing Systems 37 (2024) 12293–12333.

[15] A. Mumuni, F. Mumuni, N. K. Gerrar, A survey of synthetic data augmentation methods in machine vision, Machine Intelligence Research 21 (5) (2024) 831–869.

[16] Y. Gong, Y.-A. Chung, J. Glass, Ast: Audio spectrogram transformer, arXiv preprint arXiv:2104.01778 (2021).

[17] M. H. Erol, A. Senocak, J. Feng, J. S. Chung, Audio mamba: Bidirectional state space model for audio representation learning, IEEE Signal Processing Letters (2024).

[18] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[19] P. Warden, Speech commands: A dataset for limited-vocabulary speech recognition, arXiv preprint arXiv:1804.03209 (2018).

[20] K. J. Piczak, ESC: Dataset for Environmental Sound Classification, in: Proceedings of the 23rd Annual ACM Conference on Multimedia, ACM Press, pp. 1015–1018. doi:10.1145/2733373.2806390.
URL http://dl.acm.org/citation.cfm?doid=2733373.2806390

[21] J. Salamon, C. Jacoby, J. P. Bello, A dataset and taxonomy for urban sound research, in: 22nd ACM International Conference on Multimedia (ACM-MM'14), Orlando, FL, USA, 2014, pp. 1041–1044.

[22] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE Transactions on speech and audio processing 10 (5) (2002) 293–302.

[23] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, Q. V. Le, Specaugment: A simple data augmentation method for automatic speech recognition, in: Proc. Interspeech 2019, 2019, pp. 2613–2617.

[24] L. Graves, V. Nagisetty, V. Ganesh, Amnesiac machine learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 11516–11524.

[25] V. S. Chundawat, A. K. Tarun, M. Mandal, M. Kankanhalli, Can bad teaching induce forgetting? unlearning in deep networks using an incompetent teacher, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, 2023, pp. 7210–7217.

[26] M. Chen, W. Gao, G. Liu, K. Peng, C. Wang, Boundary unlearning: Rapid forgetting of deep networks via shifting the decision boundary, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7766–7775.

[27] A. K. Tarun, V. S. Chundawat, M. Mandal, M. Kankanhalli, Fast yet effective machine unlearning, IEEE Transactions on Neural Networks and Learning Systems 35 (9) (2023) 13046–13055.

[28] M. Kurmanji, P. Triantafillou, J. Hayes, E. Triantafillou, Towards unbounded machine unlearning, Advances in neural information processing systems 36 (2023) 1957–1987.