Supplementary Material

Daniel Brignac
University of Arizona
Tucson, Arizona
dbrignac@arizona.edu

Niels Lobo
University of Central Florida
Orlando, Florida
niels@cs.ucf.edu

Abhijit Mahalanobis
University of Arizona
Tucson, Arizona
amahalan@arizona.edu

1. Expanded Explanation of Population Strategies

Herding. It is important to note that herding maintains maximal usage of the fixed buffer size in a class balanced manner. That is, at any given time during training on any given task with s classes observed so far, the memory buffer $\mathcal{M}$ will contain $|\mathcal{M}|$ samples with $|\mathcal{M}|/s$ samples per class (up to rounding). This allows for equal probability of sampling any sample of class s from $\mathcal{M}$ when performing batch training on $\mathcal{D}_t \cup \mathcal{M}$. This differs from reservoir sampling where there is no method for maintaining a balanced buffer and an inclination to favor storage of earlier task samples leading to greater forgetting.

Another important aspect of herding is the overwriting of stored data when the fixed buffer is saturated and we wish to store newly encountered samples. In order to preserve the samples that best represent the learned class mean μ_c, we must overwrite the samples that are least informative to μ_c. As herding greedily selects mean preserving samples, the overwriting of the least informative sample corresponds to the replacement of the last sample(s) added to each class which in turn preserves the class balancing of the fixed buffer. This ensures the storage of the best samples, as determined by herding, at any given time.

GSS. When the buffer is not saturated, samples are randomly added as encountered along with their score to the buffer $\mathcal{M}$. When the buffer is full, these buffer samples are then randomly selected as candidate samples to be replaced. If a candidate's sample score is less than the score of the sample selected to be added, then that candidate is replaced in $\mathcal{M}$ by the selected sample.

IPM. Intuitively, IPM can be thought of as finding the first singular right vector of $\boldsymbol{A}_c$, where the constraint $||\boldsymbol{v}|| = 1$ keeps $\boldsymbol{v}$ on the unit sphere. We then search for the data point (in feature space) that produces the smallest angle with $\boldsymbol{v}$ and store the corresponding input sample in $\mathcal{M}$ to be used for replay. All other data points are then projected onto the null space of $\boldsymbol{\rho}$ to obtain the new matrix $\boldsymbol{A}_c(\boldsymbol{I} - \boldsymbol{\rho}_{m^{(1)}}\boldsymbol{\rho}_{m^{(1)}}^T)$ where $\boldsymbol{I}$ is the identity matrix. The process is then repeated for the K desired number of samples to be stored in $\mathcal{M}$. This ensures that each selected sample is orthogonal to previously selected samples ensuring that data in $\mathcal{M}$ is minimally redundant and overall leading to less forgetting.

2. Hyperparameter Configurations

All tested training methods use the SGD optimizer. The following hyperparameters apply to all datasets with any buffer size (both fixed and dynamic) unless stated otherwise. Note that both split-CIFAR10 and split-CIFAR100 are trained for 50 epochs and split-TinyImageNet is trained for 100 epochs for all training methods.

ER uses a learning rate of 0.1, a mini-batch size of 32, and a buffer mini-batch size of 32. DER uses a learning rate of 0.03, a mini-batch size of 32, a buffer mini-batch size of 32, an alpha of 0.3 for split-CIFAR10 and split-CIFAR100 and an alpha of 0.1 for split-TinyImageNet. GDumb uses a learning rate of 0.1 with weight decay factor of $1\mathrm{e}{-6}$, a mini-batch size of 32, a buffer mini-batch size of 32, and fits to the buffer for 250 epochs. ER-ACE uses a learning rate of 0.03, a mini-batch size of 32, and a buffer mini-batch size of 32.

3. Supplemental Results

Table 1 shows the final dynamic buffer sizes when using dynamic buffers with their respective criterion. We see that the Kaiser criterion leads to overall smaller final buffer sizes compared to intracluster variance. Figures 11 through 13 show the *FF* and *FAA* for each of the three datasets. In general, we see that the Kaiser criterion yields better performance gains when compared to the large buffer sizes of intracluster variance, particularly in reservoir sampling and IPM.

Tables 2 through 4 show task-IL results for fixed buffer and class-IL and task-IL results for dynamic buffers. We observe much of the same trends in the task-IL setting where reservoir sampling tends to lead to higher forgetting when compared to other population strategies.

Dynamic Buffer Criterion	**Split-CIFAR10**	**Split-CIFAR100**	**Split-TinyImageNet**
Kaiser Criterion	2483	4998	11849
Intracluster Variance	1000	10000	32000

Table 1: Final buffer sizes for each of the proposed dynamic buffer criterion.

Interestingly, we observe inferior performance when herding and DER are applied together with dynamic buffers. Figure 1 shows the results for the task to task performance of this pairing. We observe that for $t > 1$, there is complete forgetting of each task by the start of the next subsequent task. This would indicate that the memory buffer is potentially not being populated with subsequent task data, however, upon inspection of the buffer, each task has proper representation in the buffer as determined by each of the dynamic buffer criterion. A possible explanation for this behavior compared to the fixed buffer behavior would be maximal usage of the buffer space at all times with fixed buffers. When the buffer is dynamic, each class is never populated with more than what the dynamic buffer criterion tell us, whereas in fixed buffers, the buffer is allowed to contain a maximal amount of task data for the so far encountered tasks, up to equivalence between tasks.

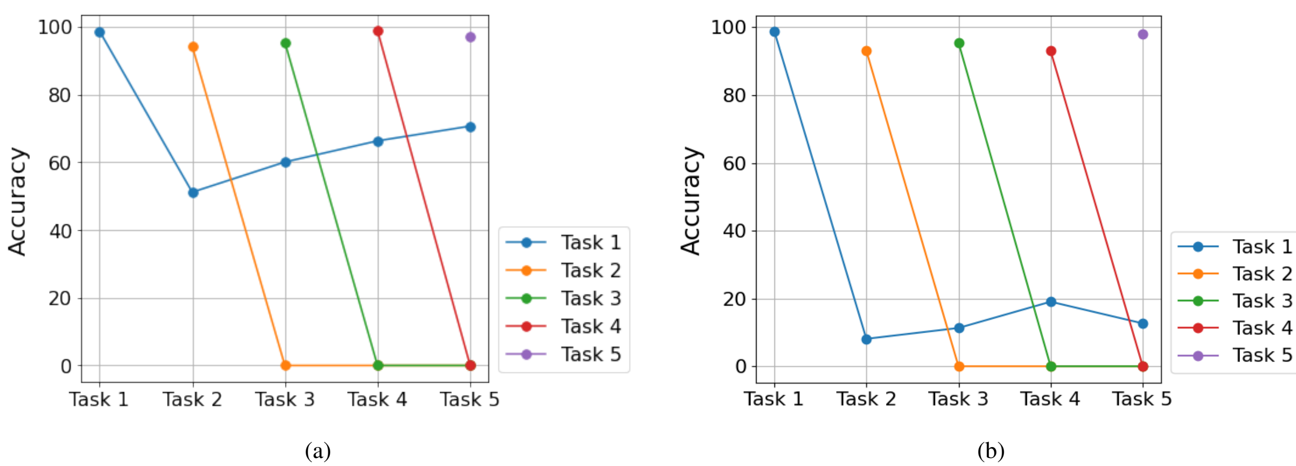


Figure 1: DER results using the herding population strategy for both (a) the Kaiser Criterion and (b) Intracluster Variance.

Fixed Buffer Size	Method	Population Strategy	Split-CIFAR10 Task-IL *FAA*	Split-CIFAR10 Task-IL *FF*	Split-CIFAR100 Task-IL *FAA*	Split-CIFAR100 Task-IL *FF*	Split-TinyImageNet Task-IL *FAA*	Split-TinyImageNet Task-IL *FF*
200	ER	Reservoir	91.65 ± 0.43	6.45 ± 0.31	65.74 ± 0.34	25.45 ± 0.24	38.47 ± 0.75	43.51 ± 0.70
		Herding	$\mathbf{92.89 \pm 0.56}$	$\mathbf{4.75 \pm 0.89}$	$\mathbf{69.71 \pm 1.04}$	$\mathbf{20.15 \pm 0.77}$	$\mathbf{44.05 \pm 0.86}$	$\mathbf{37.5 \pm 1.50}$
		GSS	89.26 ± 1.87	9.42 ± 2.21	57.68 ± 0.48	33.64 ± 0.64	-	-
		IPM	91.49 ± 0.04	6.13 ± 0.41	66.41 ± 1.23	23.77 ± 0.95	41.87 ± 0.41	40.16 ± 0.52
	DER	Reservoir	91.08 ± 0.56	7.08 ± 0.20	66.64 ± 1.53	25.57 ± 1.88	40.30 ± 1.51	42.58 ± 1.95
		Herding	90.95 ± 1.17	6.98 ± 1.51	57.61 ± 1.55	34.83 ± 1.84	19.88 ± 2.92	56.16 ± 1.56
		GSS	77.03 ± 3.82	23.81 ± 4.57	40.43 ± 5.32	50.75 ± 4.94	-	-
		IPM	$\mathbf{92.56 \pm 0.16}$	$\mathbf{4.72 \pm 0.34}$	$\mathbf{69.8 \pm 0.41}$	$\mathbf{21.65 \pm 0.29}$	$\mathbf{47.12 \pm 0.49}$	$\mathbf{33.61 \pm 0.43}$
	GDumb	Reservoir	67.95 ± 1.67	N/A	22.31 ± 1.01	N/A	11.08 ± 0.57	N/A
		Herding	$\mathbf{72.29 \pm 0.34}$	N/A	$\mathbf{28.10 \pm 1.31}$	N/A	$\mathbf{15.65 \pm 1.04}$	N/A
		GSS	66.40 ± 0.67	N/A	19.53 ± 0.38	N/A	-	N/A
		IPM	70.67 ± 0.55	N/A	26.13 ± 0.09	N/A	14.33 ± 0.42	N/A
	ER-ACE	Reservoir	93.37 ± 0.56	4.49 ± 1.20	69.21 ± 0.89	22.69 ± 0.89	44.20 ± 0.35	38.37 ± 0.31
		Herding	$\mathbf{93.81 \pm 0.27}$	$\mathbf{3.66 \pm 0.45}$	$\mathbf{73.82 \pm 0.09}$	17.75 ± 0.24	48.54 ± 0.38	32.77 ± 0.85
		GSS	79.92 ± 1.48	21.27 ± 1.77	64.79 ± 0.99	$\mathbf{15.31 \pm 9.65}$	-	-
		IPM	91.34 ± 0.71	6.90 ± 0.79	72.09 ± 0.65	19.54 ± 0.66	$\mathbf{50.94 \pm 0.52}$	$\mathbf{30.77 \pm 0.62}$
500	ER	Reservoir	93.88 ± 0.32	3.65 ± 0.67	74.1 ± 0.86	15.58 ± 0.53	49.64 ± 0.94	31.64 ± 0.82
		Herding	$\mathbf{94.48 \pm 0.41}$	$\mathbf{2.56 \pm 0.37}$	$\mathbf{76.24 \pm 1.14}$	$\mathbf{13.40 \pm 0.99}$	$\mathbf{51.33 \pm 0.75}$	$\mathbf{30.25 \pm 0.98}$
		GSS	92.21 ± 1.66	5.63 ± 2.10	61.54 ± 1.92	29.00 ± 2.48	-	-
		IPM	94.12 ± 0.33	3.50 ± 0.67	73.87 ± 0.34	15.87 ± 0.34	48.92 ± 0.12	32.09 ± 0.51
	DER	Reservoir	92.91 ± 0.13	4.96 ± 0.55	$\mathbf{74.27 \pm 0.53}$	16.88 ± 0.42	$\mathbf{52.49 \pm 0.47}$	29.72 ± 0.60
		Herding	92.92 ± 0.22	$\mathbf{3.35 \pm 1.12}$	61.17 ± 1.07	30.69 ± 1.01	18.87 ± 2.86	53.50 ± 1.61
		GSS	80.10 ± 3.31	19.10 ± 3.23	47.83 ± 3.14	45.00 ± 3.22	-	-
		IPM	$\mathbf{93.52 \pm 0.60}$	3.92 ± 1.23	74.20 ± 0.55	$\mathbf{16.84 \pm 0.40}$	51.89 ± 0.62	$\mathbf{26.70 \pm 0.48}$
	GDumb	Reservoir	78.27 ± 0.23	N/A	31.50 ± 0.93	N/A	15.62 ± 0.78	N/A
		Herding	$\mathbf{79.57 \pm 1.15}$	N/A	$\mathbf{35.41 \pm 0.78}$	N/A	$\mathbf{19.47 \pm 0.56}$	N/A
		GSS	74.77 ± 0.17	N/A	21.40 ± 2.05	N/A	-	N/A
		IPM	78.31 ± 1.29	N/A	31.85 ± 0.61	N/A	15.88 ± 0.66	N/A
	ER-ACE	Reservoir	94.16 ± 0.46	3.41 ± 0.71	76.09 ± 0.11	15.07 ± 0.52	53.55 ± 0.60	27.92 ± 1.11
		Herding	$\mathbf{95.17 \pm 0.37}$	$\mathbf{2.19 \pm 0.64}$	$\mathbf{79.23 \pm 0.50}$	$\mathbf{11.65 \pm 0.67}$	$\mathbf{56.21 \pm 0.41}$	$\mathbf{25.09 \pm 0.57}$
		GSS	79.57 ± 0.63	21.5 ± 0.68	68.98 ± 0.34	18.01 ± 6.73	-	-
		IPM	93.77 ± 0.14	3.74 ± 0.48	77.23 ± 0.39	13.68 ± 0.60	55.48 ± 0.02	26.30 ± 0.18
5120	ER	Reservoir	96.83 ± 0.29	0.58 ± 0.19	86.12 ± 0.25	3.80 ± 0.33	$\mathbf{68.01 \pm 0.17}$	$\mathbf{11.16 \pm 0.57}$
		Herding	$\mathbf{97.19 \pm 0.11}$	0.43 ± 0.17	86.23 ± 0.57	$\mathbf{3.58 \pm 0.26}$	67.00 ± 0.07	12.00 ± 0.08
		GSS	91.84 ± 3.40	6.67 ± 4.25	71.02 ± 1.39	21.30 ± 1.45	-	-
		IPM	97.08 ± 0.16	$\mathbf{0.23 \pm 0.04}$	$\mathbf{85.53 \pm 0.63}$	4.15 ± 0.46	67.04 ± 0.45	11.85 ± 0.23
	DER	Reservoir	95.38 ± 0.05	1.86 ± 0.25	$\mathbf{85.56 \pm 0.07}$	5.67 ± 0.30	$\mathbf{69.28 \pm 0.33}$	11.04 ± 0.36
		Herding	$\mathbf{96.92 \pm 0.24}$	$\mathbf{0.46 \pm 0.29}$	85.12 ± 0.28	$\mathbf{5.53 \pm 0.15}$	18.02 ± 5.35	36.75 ± 4.48
		GSS	84.15 ± 4.41	15.60 ± 5.55	56.32 ± 2.46	35.91 ± 2.64	-	-
		IPM	95.05 ± 0.50	1.25 ± 0.72	84.37 ± 0.15	5.98 ± 0.31	68.52 ± 0.58	$\mathbf{9.45 \pm 0.65}$
	GDumb	Reservoir	$\mathbf{94.85 \pm 0.49}$	N/A	71.12 ± 0.42	N/A	45.51 ± 0.42	N/A
		Herding	94.60 ± 1.48	N/A	66.91 ± 0.55	N/A	41.67 ± 0.22	N/A
		GSS	91.80 ± 0.18	N/A	51.85 ± 4.28	N/A	-	N/A
		IPM	94.48 ± 0.43	N/A	$\mathbf{71.42 \pm 0.28}$	N/A	$\mathbf{45.55 \pm 0.54}$	N/A
	ER-ACE	Reservoir	96.85 ± 0.05	0.5 ± 0.12	86.39 ± 0.12	4.07 ± 0.16	70.01 ± 0.06	9.37 ± 0.24
		Herding	$\mathbf{97.14 \pm 0.16}$	$\mathbf{0.35 \pm 0.05}$	$\mathbf{87.63 \pm 0.12}$	$\mathbf{3.03 \pm 0.02}$	69.72 ± 0.43	10.33 ± 0.48
		GSS	80.17 ± 0.53	21.00 ± 0.37	75.03 ± 0.7	16.34 ± 0.89	-	-
		IPM	96.93 ± 0.18	0.37 ± 0.04	86.91 ± 0.23	3.50 ± 0.12	$\mathbf{70.77 \pm 0.19}$	$\mathbf{9.00 \pm 0.49}$

Table 2: Task-IL population strategy results tested with various replay based methods with traditionally used fixed size buffer, averaged across three runs. We do not report forgetting in GDumb experiments due to the nature of GDumb only training on the fully populated, balanced buffer. Results for TinyImageNet are not reported for GSS due to intractable train times.

Dynamic Buffer Criterion	Method	Population Strategy	Split-CIFAR10 Class-IL *FAA*	Split-CIFAR10 Class-IL *FF*	Split-CIFAR100 Class-IL *FAA*	Split-CIFAR100 Class-IL *FF*	Split-TinyImageNet Class-IL *FAA*	Split-TinyImageNet Class-IL *FF*
Kaiser Criterion	ER	Reservoir	**75.96 ± 0.67**	25.45 ± 0.52	47.39 ± 1.15	43.89 ± 0.88	34.37 ± 0.18	45.95 ± 0.20
		Herding	75.66 ± 1.70	24.58 ± 6.24	**49.11 ± 41.48**	**41.48 ± 0.25**	**39.95 ± 0.24**	**43.71 ± 0.43**
		IPM	72.72 ± 1.06	**25.14 ± 5.20**	44.84 ± 1.33	45.12 ± 0.80	33.35 ± 0.46	46.23 ± 0.35
	DER	Reservoir	**77.46 ± 0.58**	22.33 ± 0.80	**57.27 ± 0.49**	30.44 ± 1.21	**41.53 ± 0.54**	29.40 ± 1.48
		Herding	27.22 ± 7.55	84.56 ± 6.24	10.17 ± 0.20	87.52 ± 0.38	7.42 ± 0.67	69.22 ± 3.57
		IPM	66.86 ± 1.45	**13.02 ± 0.57**	53.64 ± 0.38	**12.64 ± 0.56**	34.69 ± 0.17	**8.54 ± 0.29**
	GDumb	Reservoir	69.81 ± 1.07	N/A	40.87 ± 0.63	N/A	**31.30 ± 0.44**	N/A
		Herding	65.75 ± 0.7	N/A	36.89 ± 0.65	N/A	27.74 ± 046	N/A
		IPM	**71.65 ± 0.48**	N/A	**41.03 ± 0.29**	N/A	30.02 ± 0.40	N/A
	ER-ACE	Reservoir	**79.53 ± 0.76**	19.01 ± 1.08	**53.23 ± 0.47**	35.10 ± 0.69	39.48 ± 0.32	38.72 ± 0.77
		Herding	69.22 ± 0.67	12.47 ± 0.74	51.91 ± 0.10	**13.4 ± 0.19**	**41.65 ± 1.75**	**14.20 ± 0.07**
		IPM	61.49 ± 0.28	**9.77 ± 1.02**	49.98 ± 0.12	16.55 ± 0.25	38.99 ± 0.34	16.43 ± 0.17
Intracluster Variance	ER	Reservoir	65.30 ± 0.33	39.22 ± 0.38	56.07 ± 0.28	32.62 ± 0.28	**45.40 ± 0.35**	29.90 ± 0.32
		Herding	**67.98 ± 0.44**	**35.51 ± 0.72**	**56.19 ± 0.85**	**30.98 ± 0.43**	44.91 ± 0.39	**29.07 ± 0.45**
		IPM	65.79 ± 0.45	36.75 ± 2.17	53.53 ± 1.30	33.66 ± 0.39	44.12 ± 0.51	30.62 ± 0.14
	DER	Reservoir	**69.27 ± 0.59**	33.22 ± 1.01	**62.16 ± 0.14**	22.55 ± 0.52	42.30 ± 0.33	25.47 ± 0.55
		Herding	22.47 ± 0.24	92.09 ± 0.07	10.16 ± 0.03	88.01 ± 0.20	8.04 ± 0.07	76.55 ± 0.20
		IPM	65.49 ± 0.21	**16.53 ± 1.39**	57.32 ± 0.57	**8.07 ± 1.71**	**33.47 ± 1.10**	**10.63 ± 3.95**
	GDumb	Reservoir	49.73 ± 3.59	N/A	**54.44 ± 0.42**	N/A	46.30 ± 0.35	N/A
		Herding	42.73 ± 3.49	N/A	48.95 ± 0.87	N/A	42.96 ± 0.42	N/A
		IPM	**51.94 ± 2.23**	N/A	54.27 ± 0.76	N/A	**46.37 ± 0.44**	N/A
	ER-ACE	Reservoir	**74.34 ± 1.29**	26.57 ± 1.84	**59.69 ± 0.24**	24.74 ± 0.23	**47.00 ± 0.05**	25.66 ± 0.21
		Herding	61.62 ± 0.64	14.46 ± 1.59	56.29 ± 0.32	**10.69 ± 0.39**	45.90 ± 0.06	**11.15 ± 0.06**
		IPM	56.40 ± 0.59	**14.38 ± 0.91**	56.05 ± 0.77	14.06 ± 1.10	43.60 ± 0.21	13.37 ± 0.67

Table 3: Class-IL population strategy results tested with various replay based methods with the proposed dynamic buffer criterion, averaged across three runs. We do not report forgetting in GDumb experiments due to the nature of GDumb only training on the fully populated, balanced buffer. GSS experiments are omitted due to inferior performance with fixed buffer sizes.

Dynamic Buffer Criterion	Method	Population Strategy	Split-CIFAR10 Task-IL *FAA*	Split-CIFAR10 Task-IL *FF*	Split-CIFAR100 Task-IL *FAA*	Split-CIFAR100 Task-IL *FF*	Split-TinyImageNet Task-IL *FAA*	Split-TinyImageNet Task-IL *FF*
Kaiser Criterion	ER	Reservoir	**95.60 ± 0.10**	1.77 ± 0.04	83.77 ± 1.13	5.93 ± 0.82	70.16 ± 0.23	8.50 ± 0.51
		Herding	95.50 ± 0.67	1.44 ± 0.64	**84.60 ± 0.48**	**4.75 ± 0.29**	**71.16 ± 0.19**	**5.58 ± 2.37**
		IPM	95.39 ± 0.27	**1.24 ± 0.22**	82.33 ± 0.56	6.12 ± 0.11	68.66 ± 0.58	9.4 ± 0.23
	DER	Reservoir	91.76 ± 0.77	6.11 ± 1.39	81.33 ± 0.36	10.05 ± 0.77	**70.77 ± 0.64**	8.89 ± 0.55
		Herding	85.71 ± 6.40	11.63 ± 4.81	34.32 ± 2.33	60.72 ± 2.54	19.53 ± 2.12	56.23 ± 1.86
		IPM	**94.58 ± 0.53**	**1.68 ± 0.45**	**81.35 ± 0.23**	**8.93 ± 0.64**	69.45 ± 0.35	**7.72 ± 0.61**
	GDumb	Reservoir	91.33 ± 0.61	N/A	70.02 ± 0.37	N/A	**57.17 ± 0.40**	N/A
		Herding	89.00 ± 0.42	N/A	65.80 ± 0.88	N/A	53.11 ± 0.16	N/A
		IPM	**92.22 ± 0.54**	N/A	**70.51 ± 0.28**	N/A	55.82 ± 0.35	N/A
	ER-ACE	Reservoir	**95.81 ± 0.18**	1.75 ± 0.21	85.31 ± 0.16	5.52 ± 0.18	71.24 ± 0.28	8.45 ± 0.53
		Herding	95.81 ± 0.25	**1.37 ± 0.17**	**86.17 ± 0.25**	**3.89 ± 0.28**	**73.40 ± 0.04**	**5.54 ± 0.47**
		IPM	95.43 ± 0.10	1.95 ± 0.32	84.97 ± 0.26	5.31 ± 0.34	71.86 ± 0.22	7.23 ± 0.21
Intracluster Variance	ER	Reservoir	94.29 ± 0.16	3.42 ± 0.24	**86.69 ± 0.35**	3.06 ± 0.30	74.2 ± 0.33	4.41 ± 0.23
		Herding	**94.32 ± 0.15**	**3.00 ± 0.28**	86.57 ± 0.65	**2.25 ± 0.20**	**74.48 ± 0.28**	**2.99 ± 0.27**
		IPM	93.74 ± 0.22	3.54 ± 0.45	85.12 ± 0.78	3.11 ± 0.07	73.72 ± 0.38	3.98 ± 0.34
	DER	Reservoir	89.47 ± 0.92	9.14 ± 1.44	**85.36 ± 0.27**	6.09 ± 0.24	**76.00 ± 0.27**	4.62 ± 0.43
		Herding	79.58 ± 2.92	20.71 ± 3.36	35.95 ± 1.54	59.42 ± 1.90	20.30 ± 0.61	63.05 ± 0.79
		IPM	**93.75 ± 0.26**	**2.54 ± 0.61**	84.78 ± 0.47	**5.53 ± 0.58**	73.72 ± 0.38	**3.98 ± 0.34**
	GDumb	Reservoir	**83.28 ± 1.37**	N/A	**80.99 ± 0.17**	N/A	71.14 ± 0.40	N/A
		Herding	79.05 ± 2.35	N/A	76.66 ± 0.42	N/A	68.40 ± 0.53	N/A
		IPM	82.94 ± 2.28	N/A	80.61 ± 0.49	N/A	**73.72 ± 0.69**	N/A
	ER-ACE	Reservoir	**94.85 ± 0.24**	**2.79 ± 0.46**	87.56 ± 0.04	2.72 ± 0.18	75.35 ± 0.36	4.13 ± 0.28
		Herding	93.89 ± 0.45	3.98 ± 0.37	88.07 ± 0.14	**1.61 ± 0.21**	**76.51 ± 0.25**	**2.45 ± 0.19**
		IPM	94.01 ± 0.32	3.58 ± 0.27	**92.00 ± 3.10**	2.96 ± 0.69	75.33 ± 0.01	2.85 ± 0.25

Table 4: Task-IL population strategy results tested with various replay based methods with the proposed dynamic buffer criterion, averaged across three runs. We do not report forgetting in GDumb experiments due to the nature of GDumb only training on the fully populated, balanced buffer. GSS experiments are omitted due to inferior performance with fixed buffer sizes.

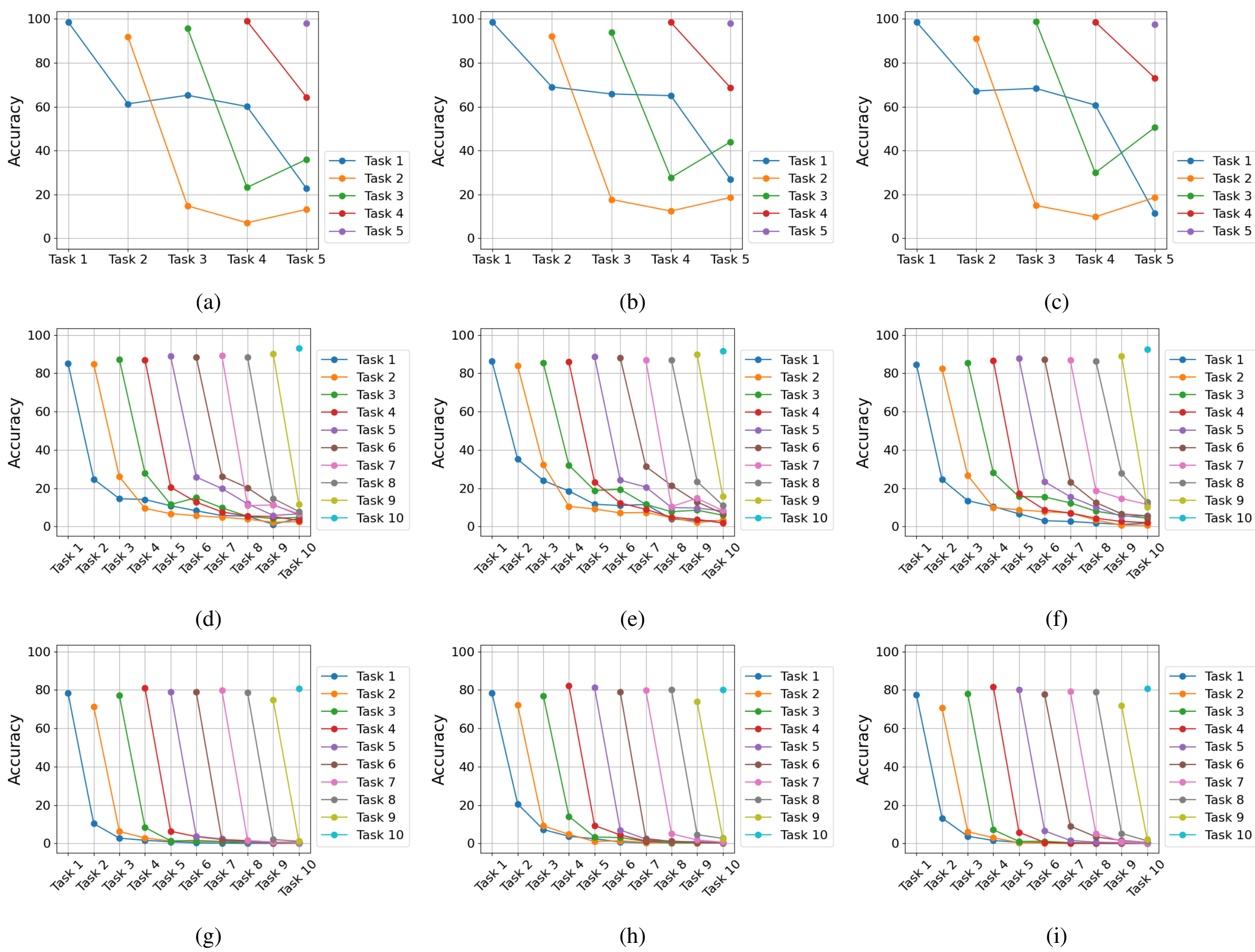


Figure 2: A comparison of the reservoir, herding, and IPM population strategies paired with ER with a fixed buffer size of 200. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

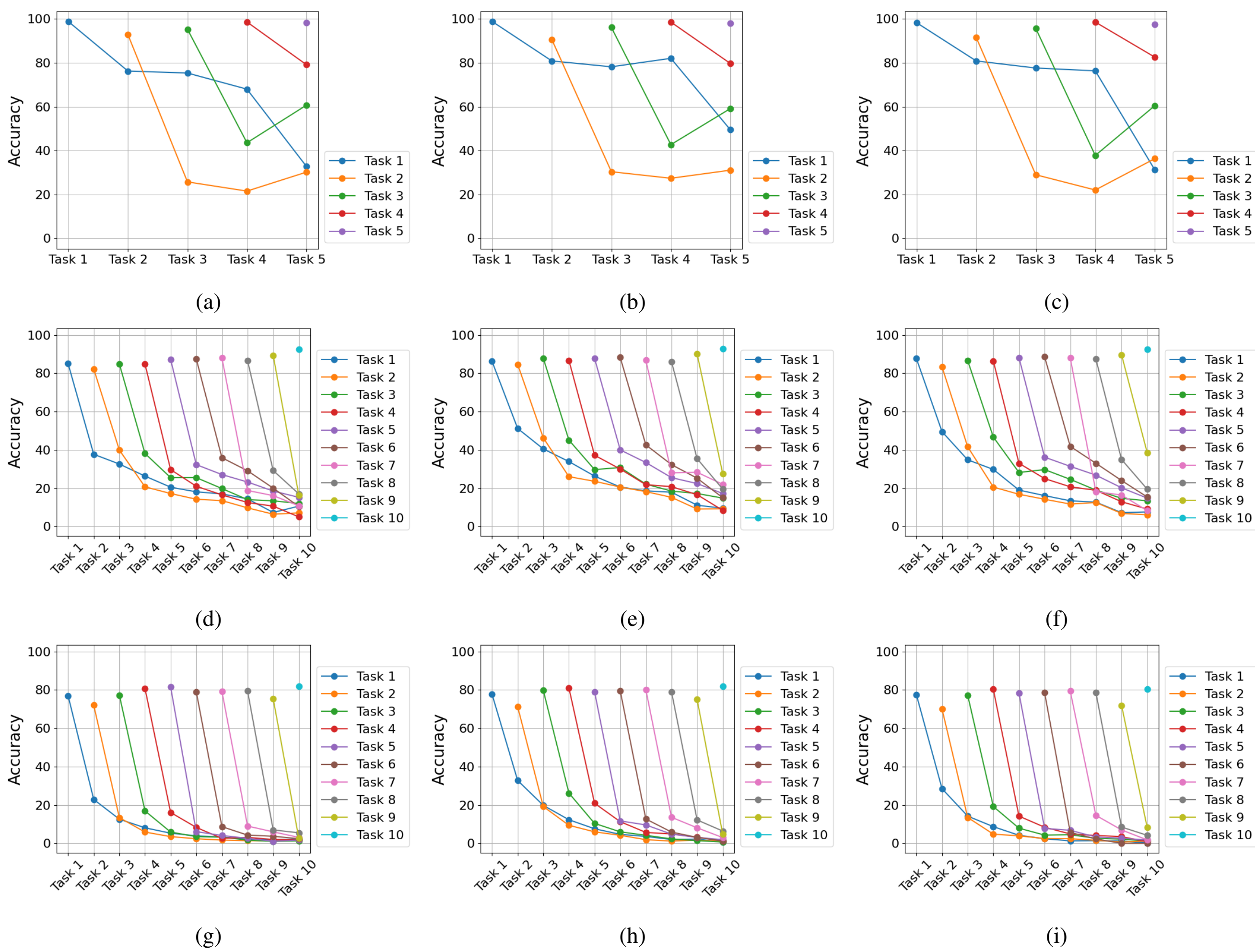


Figure 3: A comparison of the reservoir, herding, and IPM population strategies paired with ER with a fixed buffer size of 500. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

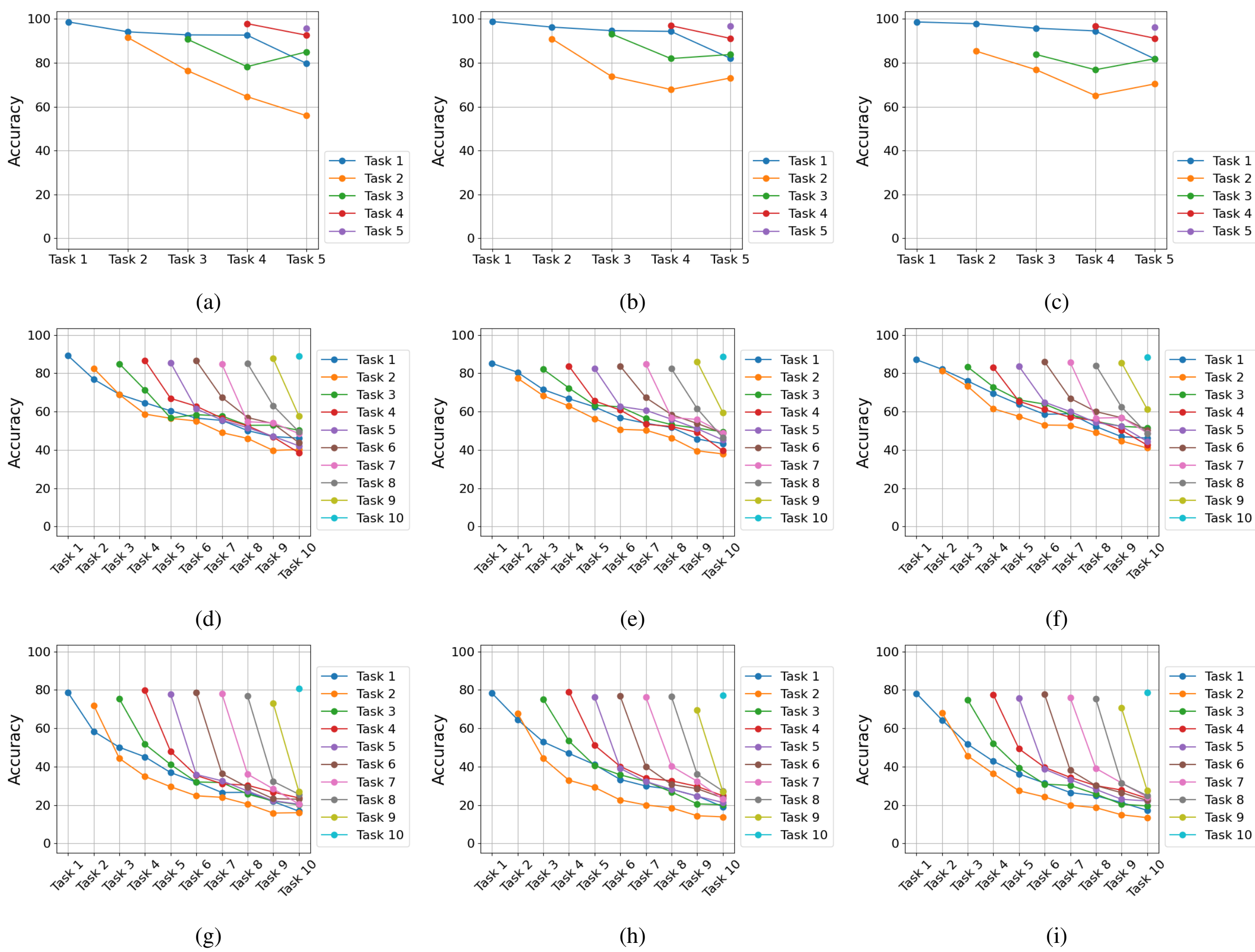


Figure 4: A comparison of the reservoir, herding, and IPM population strategies paired with ER with a fixed buffer size of 5120. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

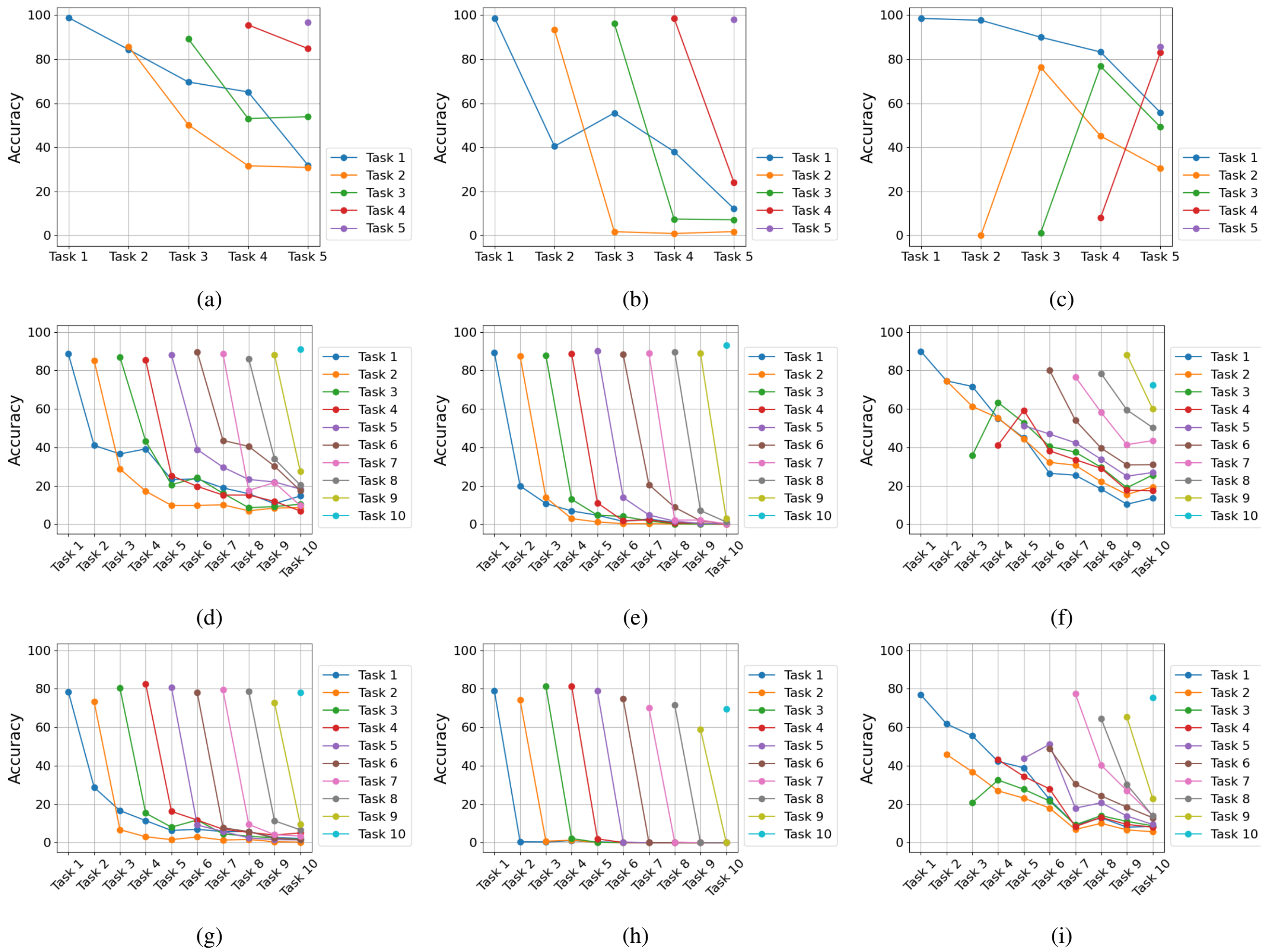


Figure 5: A comparison of the reservoir, herding, and IPM population strategies paired with DER with a fixed buffer size of 200. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

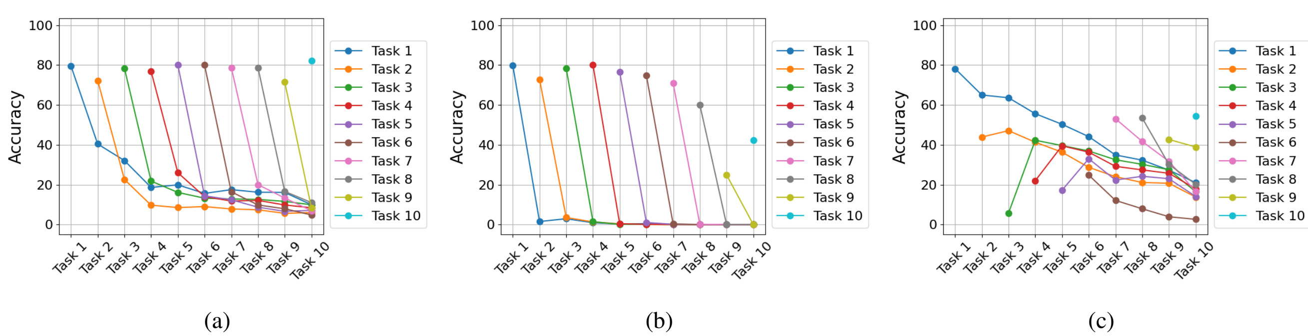


Figure 6: A comparison of the reservoir, herding, and IPM population strategies paired with DER with a fixed buffer size of 500 using the Split-TinyImageNet dataset. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

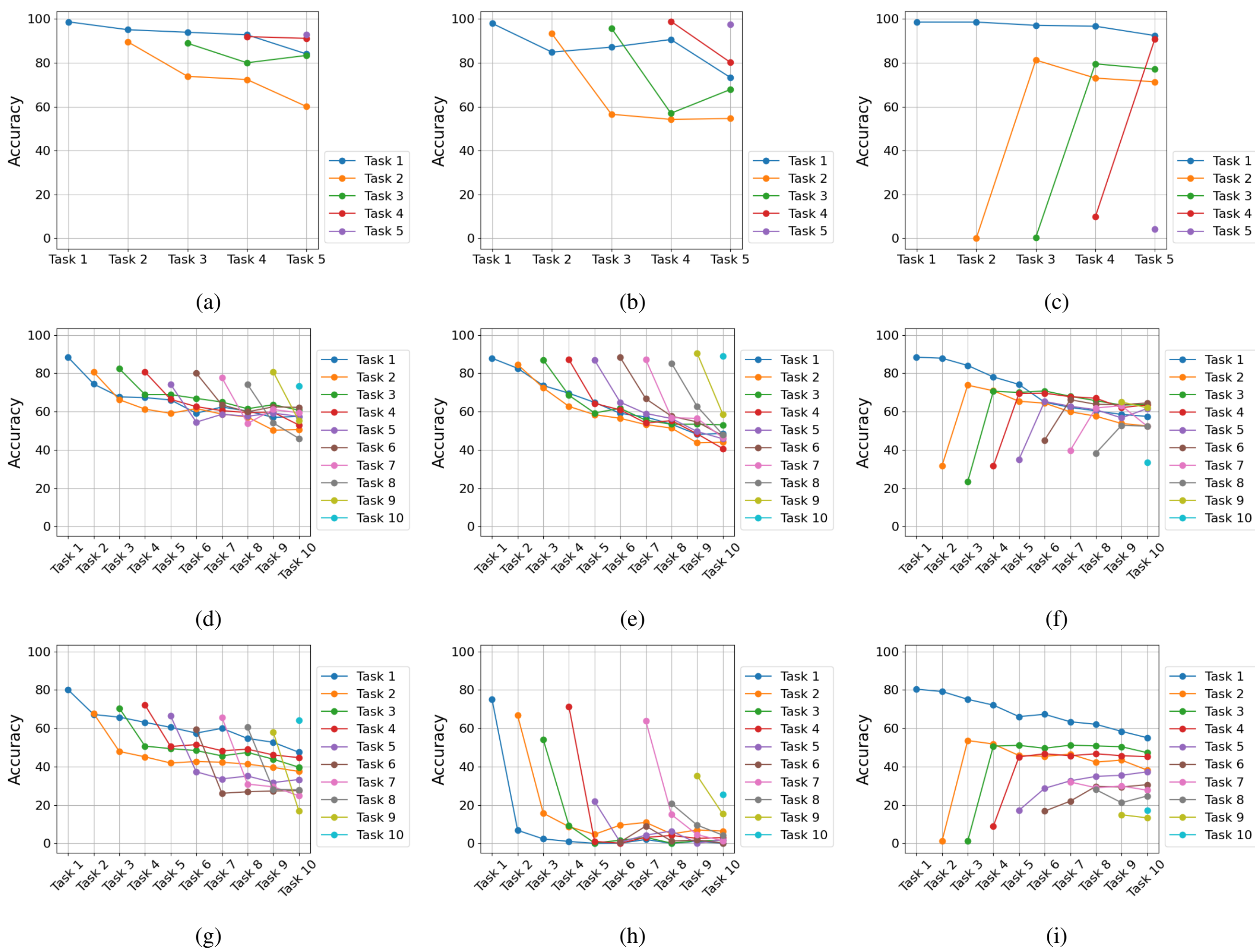


Figure 7: A comparison of the reservoir, herding, and IPM population strategies paired with DER with a fixed buffer size of 5120. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

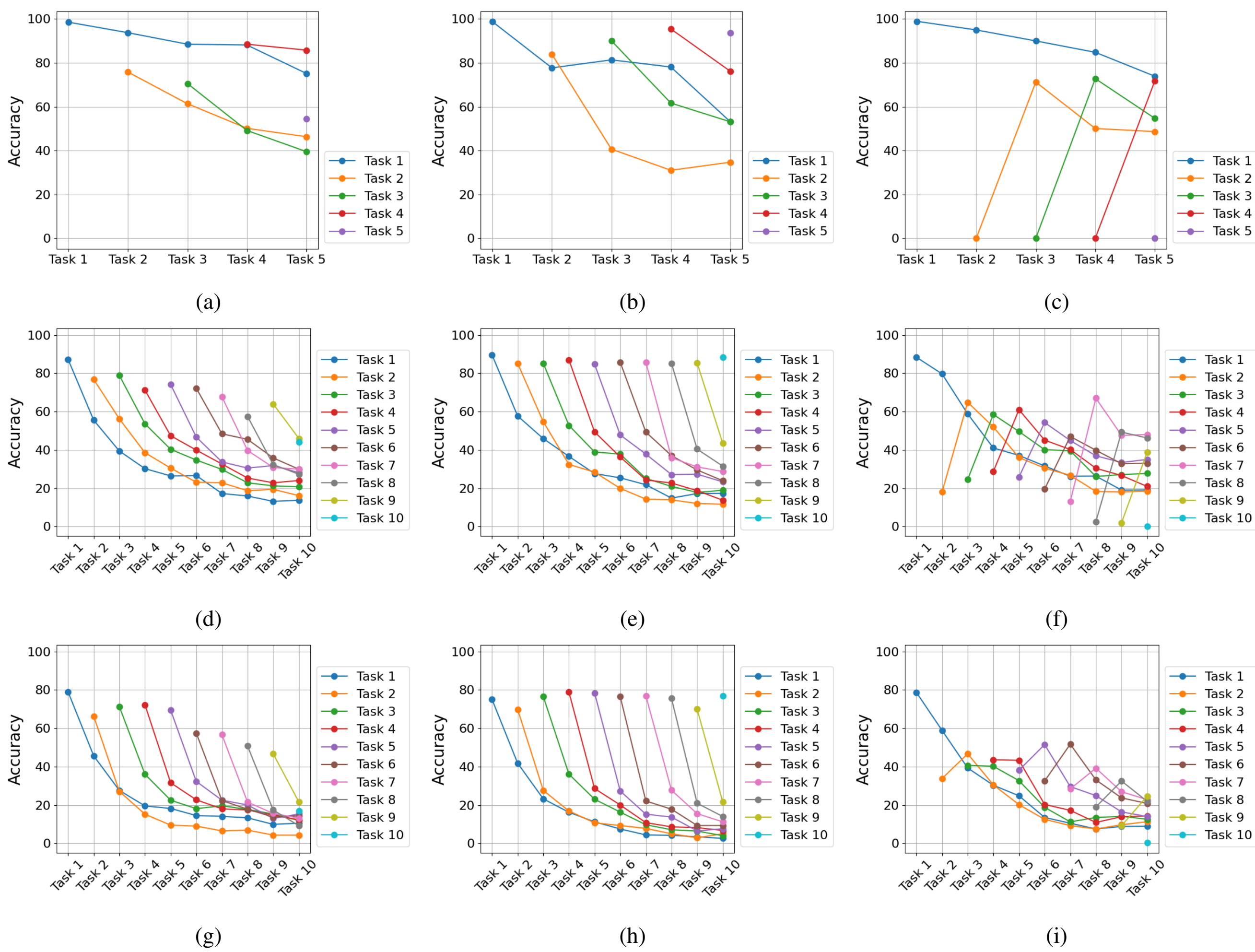


Figure 8: A comparison of the reservoir, herding, and IPM population strategies paired with ER-ACE with a fixed buffer size of 200. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

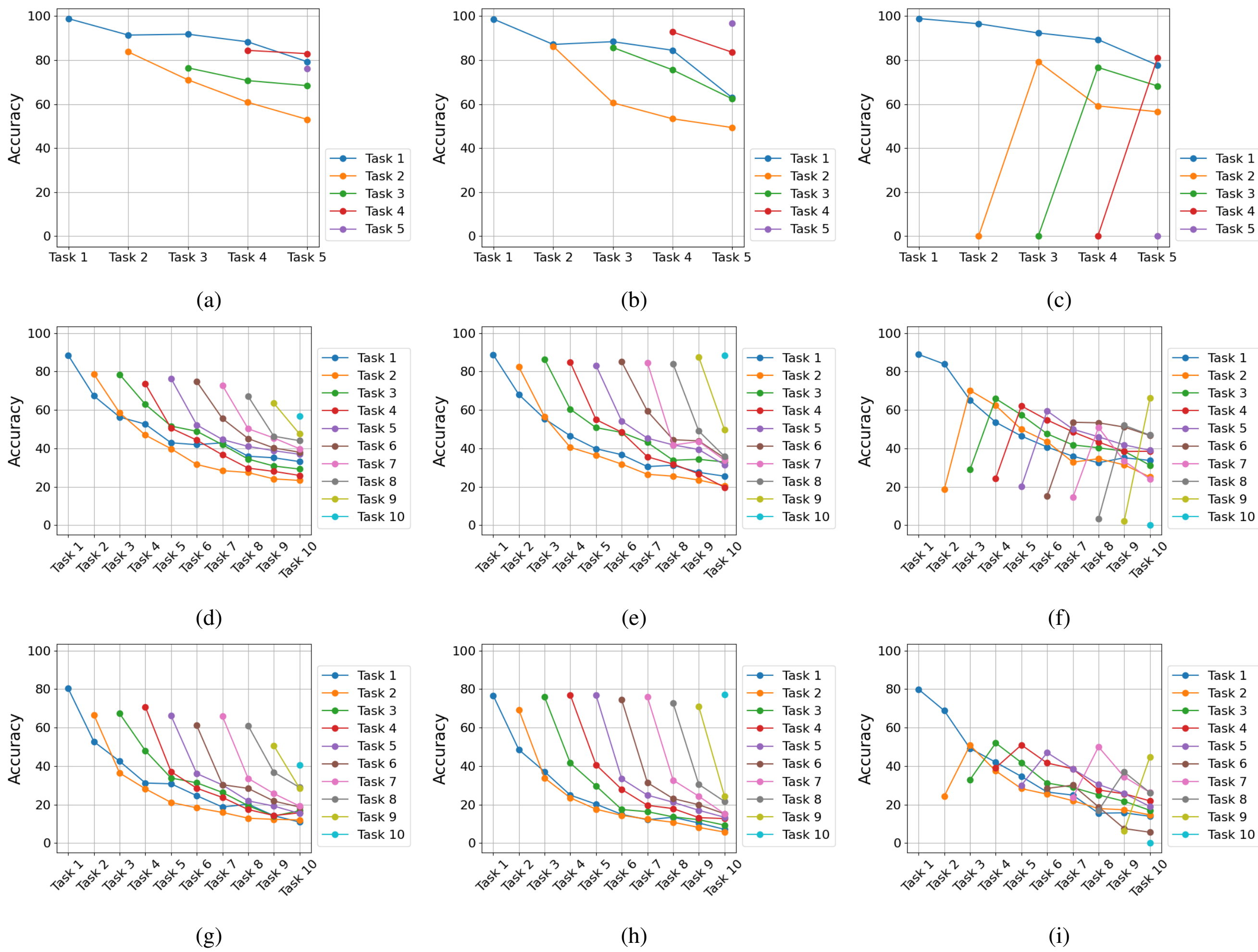


Figure 9: A comparison of the reservoir, herding, and IPM population strategies paired with ER-ACE with a fixed buffer size of 500. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

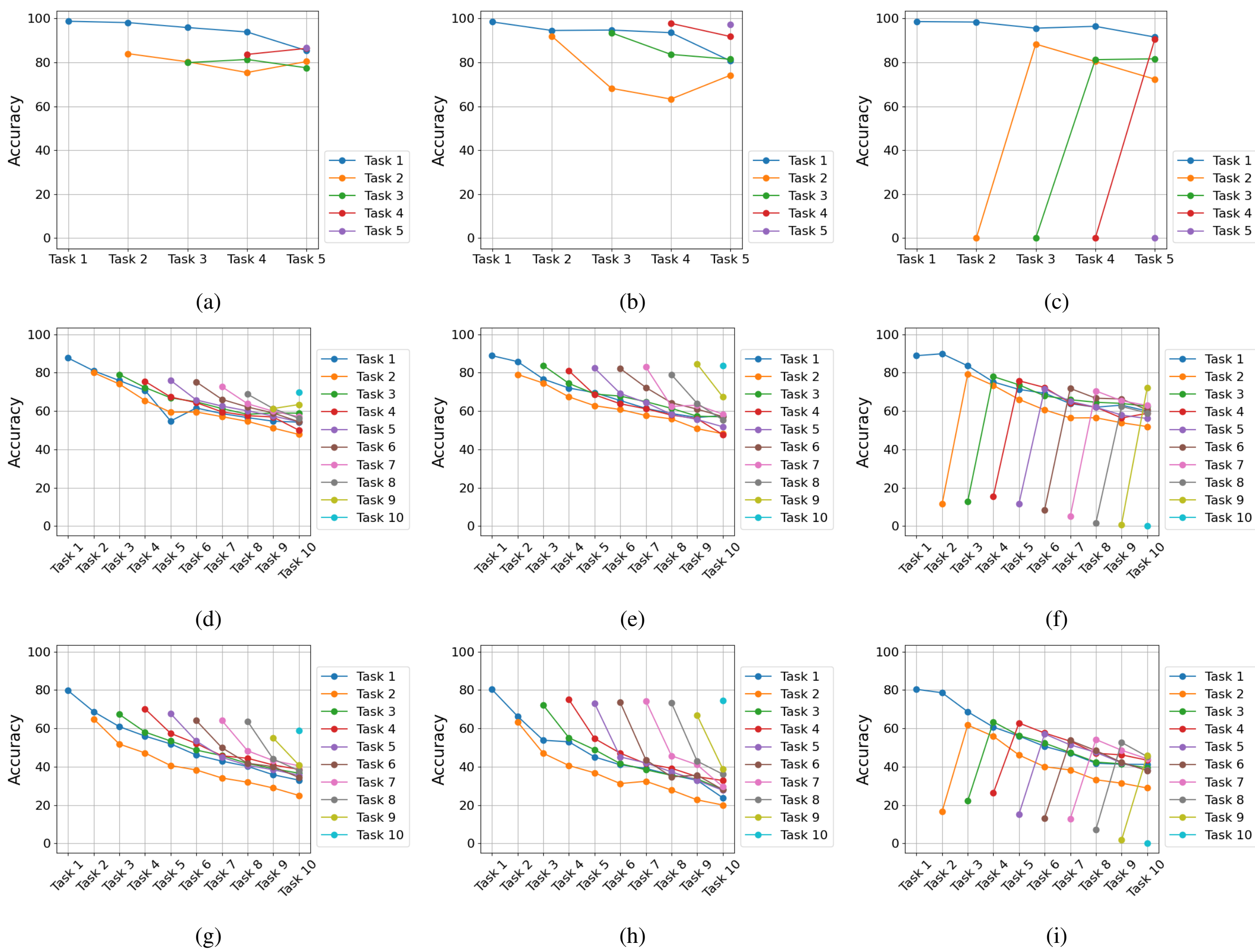


Figure 10: A comparison of the reservoir, herding, and IPM population strategies paired with ER-ACE with a fixed buffer size of 5120. Top row corresponds to Split-CIFAR10 performance, middle row is Split-CIFAR100, and bottom row is Split-TinyImageNet. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

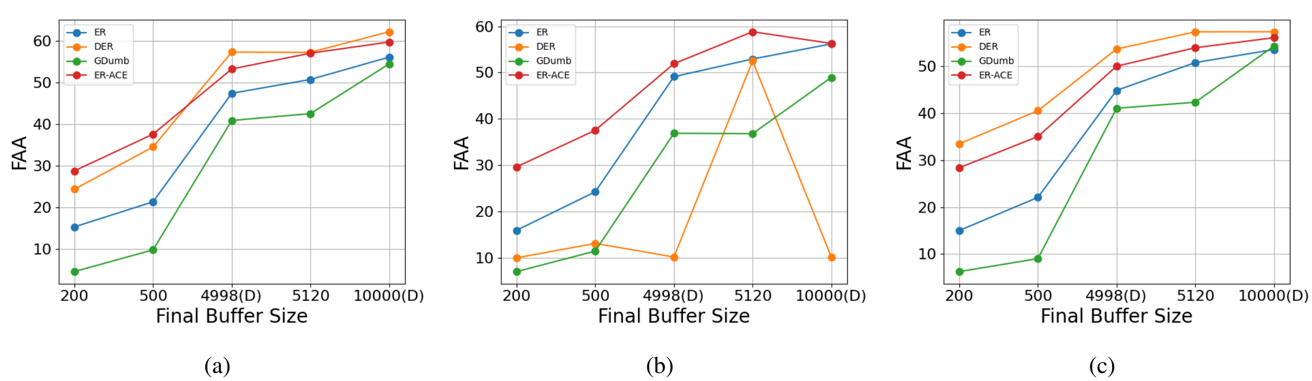
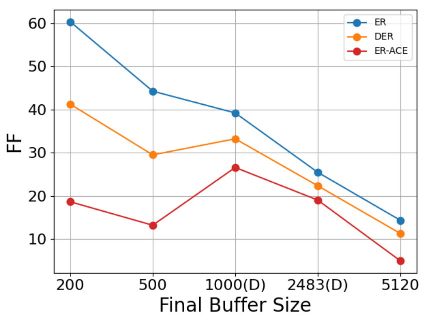
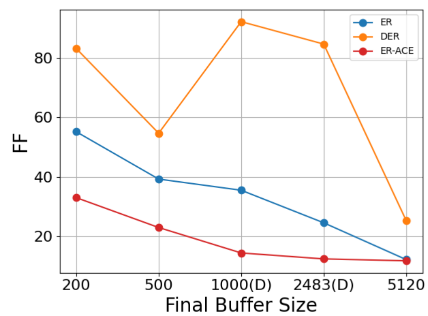


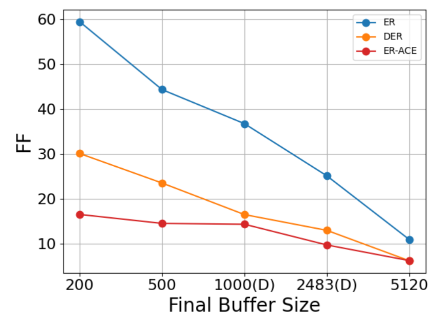
Figure 11: Final average accuracy performance with various final buffer sizes tested with Split-CIFAR100. Final buffer sizes with a (D) indicate dynamic final size. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.



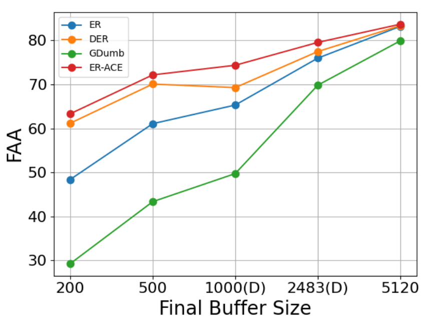
(a)



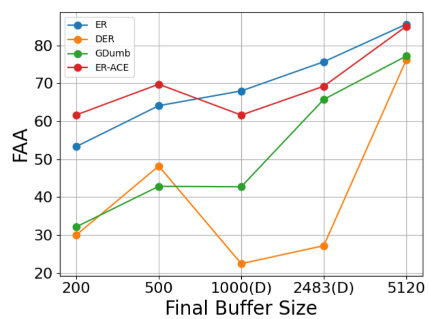
(b)



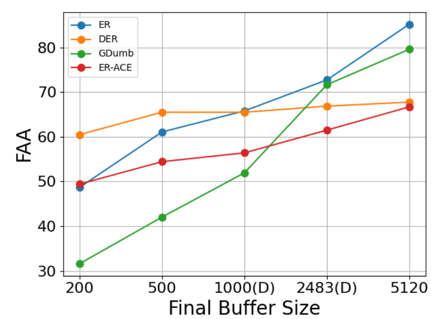
(c)



(d)



(e)



(f)

Figure 12: Final average accuracy performance with various final buffer sizes tested with Split-CIFAR10. Final buffer sizes with a (D) indicate dynamic final size. Top row shows *FF* performance while bottom row shows *FAA* performance. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.

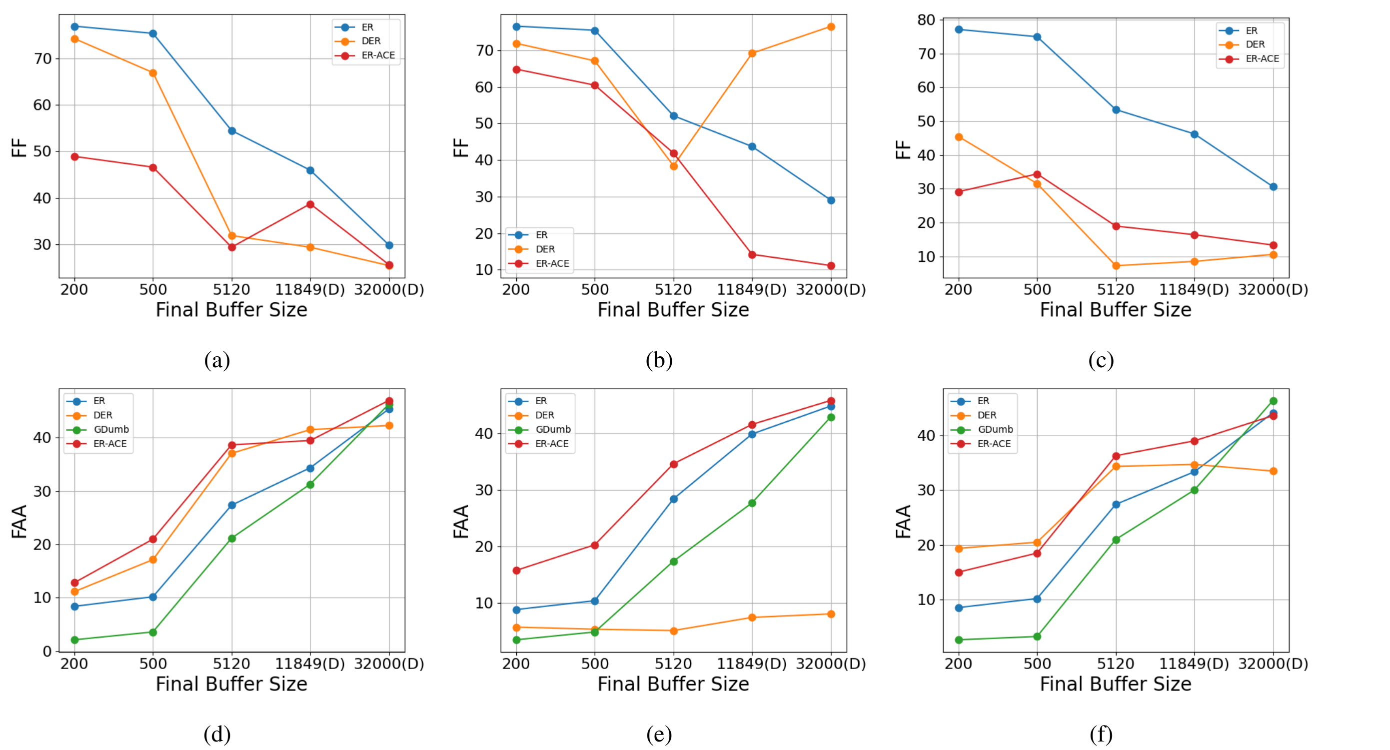


Figure 13: Final average accuracy performance with various final buffer sizes tested with Split-TinyImageNet. Final buffer sizes with a (D) indicate dynamic final size. Top row shows *FF* performance while bottom row shows *FAA* performance. The columns correspond as follows: left uses reservoir sampling, center uses herding, and right uses IPM. We do not report GSS results due to all around inferior performance.