Supplementary Material
Data-Efficient Training of CNNs and
Transformers with Coresets:
A Stability Perspective

Animesh Gupta[1], Irtiza Hasan[2], Arnav Chavan[3], Dilip K. Prasad[4], and
Deepak K. Gupta[1]

[1] Transmute AI Lab (Texmin), IIT (ISM) Dhanbad, India
[2] Immersal, Helsinki, Finland
[3] Nyun AI, India
[4] UiT The Arctic University of Norway, Tromso, Norway

A   Appendix

A.1   Introduction

We have also provided quantitative results, further illustrating the observations and analysis presented in the main manuscript.

An improved version of the code will be made publicly available with the final version of the paper.

A.2   Exploring Coreset selection Methods in Detail

CRAIG. CRAIG aims to select an optimal subset $\mathcal{S}^*$ by solving the following equation:

$$\mathcal{S}^* = \arg\min_{\mathcal{S} \subseteq \mathcal{D}, \gamma_j \geq 0 \ \forall j} |\mathcal{S}|, \ \ \text{s.t.}$$
$$\max_{\mathbf{w} \in \mathcal{W}} \| \sum_{i \in \mathcal{D}} \nabla f_i(\mathbf{w}) - \sum_{j \in \mathcal{S}} \gamma_j \nabla f_j(\mathbf{w}) \| \leq \mathcal{M}. \tag{1}$$

In the equation above, $\mathcal{D}$ denotes the training dataset, and $\mathcal{S}$ represents the selected subset with a budget $\mathcal{K}$. The optimization problem aims to find the smallest subset $\mathcal{S} \subseteq \mathcal{D}$ along with per-element stepsizes $\gamma_j > 0$ that approximate the full gradient with an error no greater than $\mathcal{M}$ for all possible optimization parameters $\mathbf{w} \in \mathcal{W}$. It is important to note that the CRAIG method can be utilized to select data subsets at different epochs.

GradMatch. We further explain GradMatch in a detailed step-by-step breakdown of the optimization problem formulation, the utilization of the Orthogonal Matching Pursuit (OMP) algorithm, and additional details regarding the computation of the error term used in GradMatch. Let $\mathcal{S}^t$ represent subsets for $t = 1, \cdots, T$, with p denoting the data points from the training or validation set.

The loss function is defined as $\mathcal{L}(\mathbf{w}) = \sum_{i \in \mathcal{W}} \mathcal{L}(\mathbf{x_i}, \mathbf{y_i}, \mathbf{w})$, where $\mathcal{L} = \mathcal{L}_\mathcal{T}$ when $\mathcal{W} = \mathcal{D}$ (training set) and $\mathcal{L} = \mathcal{L}_\mathcal{V}$ when $\mathcal{W} = \mathcal{V}$ (validation set).

$$\mathrm{Err}(\boldsymbol{\phi}^t, \mathcal{S}^t, \mathcal{L}, \mathcal{L}_\mathcal{T}, \mathbf{w_t}) = \left\| \sum_{i \in \mathcal{S}^t} \boldsymbol{\phi} \nabla_\mathbf{w} \mathcal{L}_\mathcal{T}^i(\mathbf{w_t}) - \nabla_\mathbf{w} \mathcal{L}(\mathbf{w_t}) \right\| \tag{2}$$

The error term $\mathrm{Err}(\boldsymbol{\phi}^t, \mathcal{S}^t, \mathcal{L}, \mathcal{L}_\mathcal{T}, \mathbf{w_t})$ is computed using Equation (2), where $\boldsymbol{\phi}$ is a weight vector containing the weights for each data instance, $\mathcal{K}$ represents the budget for the subset, and the L2 norm is applied. The objective is to minimize this error by solving the optimization problem using the Orthogonal Matching Pursuit (OMP) algorithm.

GLISTER. It is an approach that tackles the challenge of selecting a representative subset from the training data while optimizing model performance and efficiency. In GLISTER, the selection process involves maximizing a specific objective function, taking into account the log-likelihood on the validation set and utilizing gradient information for subset evaluation. GLISTER follows below approach to solve the optimization problem:

$$\mathcal{S}^{t+1} = \operatorname*{argmax}_{\mathcal{S} \subseteq \mathcal{D}, |\mathcal{S}| \leq \mathcal{K}} G_{\mathbf{w}^t}(\mathcal{S}), \text{where} \tag{3}$$

$$G_\mathbf{w}(\mathcal{S}) = \mathcal{LL}_\mathcal{V}(\mathbf{w} + \eta \nabla_\mathbf{w} \mathcal{LL}_\mathcal{D}(\mathbf{w}, \mathcal{S}), \mathcal{V}),$$

In the above equations, $\mathcal{D}$ represents the training set, $\mathcal{V}$ refers to the validation set, and $\mathcal{LL}_\mathcal{V}$ represents the log-likelihood on the validation set. The step size is denoted by η, and $\mathcal{K}$ represents the budget for the subset. GLISTER solves a bi-level optimization problem, where the first problem is addressed by employing an online one-step meta approximation through a single gradient step. The second approximation is achieved using the Taylor-Series approximation, implemented via the greedy search algorithm.

A.3   Results

CRAIG Method Results on CIFAR10 We initially started our experiments on the CIFAR10 dataset. Tables [A1, A2, A3, A4] gives a comparative study for these models. We used Fig. 3 and Fig. 4 to demonstrate that CNN underperforms across all coreset sizes than transformers under the pretraining setting. We can also see that at tiny coreset sizes, transformers are more stable than CNNs for the CIFAR10 dataset.

CRAIG, GradMatch, GLISTER and Random Results on Tiny ImageNet This section shows results with different subsets methods on Tiny Imagenet with ViT_B16, ResNet50, and MobileNet.

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105		147.46			-			99.01	
100	89		125.88			-			98.61	
	79		112.06			-			98.86	
	105	240.75	227.10	217.38	31.20	15.41	6.24	98.89	98.94	98.95
90	89	203.89	193.08	183.40	24.96	12.33	3.12	98.86	98.89	98.87
	79	180.12	169.14	162.60	21.84	9.25	3.12	98.73	98.78	98.78
	105	217.94	201.95	193.10	31.20	15.75	6.25	98.86	98.98	98.94
80	89	184.33	171.51	162.58	24.96	12.60	3.12	98.79	98.94	98.87
	79	162.73	149.93	144.09	21.84	9.45	3.12	98.80	98.84	98.80
	105	196.18	181.35	170.78	30.94	15.75	6.35	98.94	98.82	98.82
70	89	165.93	154.08	143.66	24.75	12.60	3.17	98.89	98.72	98.79
	79	146.63	134.69	127.53	21.66	9.45	3.17	98.92	98.78	98.69
	105	171.45	156.75	147.65	30.96	15.81	6.30	98.99	98.94	98.98
60	89	144.73	132.99	123.84	24.76	12.65	3.15	98.96	98.93	98.92
	79	127.79	115.94	109.91	21.67	9.48	3.15	98.89	98.80	98.84
	105	148.45	132.09	122.60	31.58	15.67	6.28	98.92	98.97	98.76
50	89	124.98	111.86	102.38	25.26	12.53	3.14	98.92	98.90	98.67
	79	110.22	97.18	90.84	22.10	9.40	3.14	98.81	98.86	98.71
	105	125.13	109.69	100.47	30.80	15.53	6.33	98.87	98.73	98.96
40	89	105.26	92.90	83.63	24.64	12.42	3.16	98.85	98.65	98.92
	79	92.96	80.60	74.43	21.56	9.32	3.16	98.78	98.59	98.82
	105	99.90	87.05	78.51	29.97	16.96	8.02	98.69	98.67	98.59
30	89	83.68	73.41	64.20	23.97	13.57	4.01	98.65	98.56	98.49
	79	73.80	63.11	57.25	20.97	10.17	4.01	98.57	98.58	98.46
	105	75.67	60.52	52.25	30.13	15.07	6.05	98.64	98.58	98.63
20	89	62.91	50.78	42.39	24.10	12.05	3.02	98.52	98.58	98.55
	79	55.33	43.21	37.76	21.09	9.04	3.02	98.60	98.38	98.57
	105	54.29	39.01	29.24	30.07	14.95	5.57	98.25	98.28	98.25
10	89	44.81	32.59	23.07	24.06	11.96	2.78	98.22	98.21	98.18
	79	39.50	27.30	20.82	21.05	8.97	2.78	98.25	98.28	98.25
	105	30.38	16.49	8.02	27.98	14.09	5.62	97.49	97.59	97.39
1	89	24.44	13.32	4.86	22.39	11.27	2.81	97.49	97.59	97.39
	79	21.42	10.28	4.64	19.59	8.45	2.81	97.49	97.59	97.39

Table A1: Performance scores for ViT-B_16 with ImageNet 21K weights on CIFAR10 and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		147.46			-			66.84	
100	89		125.88			-			62.21	
	79		112.06			-			64.15	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	240.75	227.10	217.38	31.20	15.41	6.24	66.66	66.76	67.03
90	89	203.89	193.08	183.40	24.96	12.33	3.12	66.42	66.46	66.85
	79	180.12	169.14	162.60	21.84	9.25	3.12	65.47	66.16	66.29
	105	217.94	201.95	193.10	31.20	15.75	6.25	65.42	65.82	66.44
80	89	184.33	171.51	162.58	24.96	12.60	3.12	65.31	65.16	66.08
	79	162.73	149.93	144.09	21.84	9.45	3.12	64.02	64.12	65.34
	105	196.18	181.35	170.78	30.94	15.75	6.35	63.77	63.53	64.88
70	89	165.93	154.08	143.66	24.75	12.60	3.17	63.28	63.52	64.59
	79	146.63	134.69	127.53	21.66	9.45	3.17	62.03	62.73	63.56
	105	171.45	156.75	147.65	30.96	15.81	6.30	61.70	62.56	63.12
60	89	144.73	132.99	123.84	24.76	12.65	3.15	61.21	61.93	62.60
	79	127.79	115.94	109.91	21.67	9.48	3.15	60.83	61.30	61.65
	105	148.45	132.09	122.60	31.58	15.67	6.28	61.15	61.44	62.02
50	89	124.98	111.86	102.38	25.26	12.53	3.14	60.78	61.07	61.99
	79	110.22	97.18	90.84	22.10	9.40	3.14	59.71	59.87	61.22
	105	125.13	109.69	100.47	30.80	15.53	6.33	59.28	60.17	60.10
40	89	105.26	92.90	83.63	24.64	12.42	3.16	58.99	59.70	59.39
	79	92.96	80.60	74.43	21.56	9.32	3.16	58.34	59.13	59.10
	105	99.90	87.05	78.51	29.97	16.96	8.02	57.75	57.67	57.79
30	89	83.68	73.41	64.20	23.97	13.57	4.01	57.15	57.27	57.53
	79	73.80	63.11	57.25	20.97	10.17	4.01	56.15	56.69	57.52
	105	75.67	60.52	52.25	30.13	15.07	6.05	52.26	55.58	53.70
20	89	62.91	50.78	42.39	24.10	12.05	3.02	49.50	55.56	52.49
	79	55.33	43.21	37.76	21.09	9.04	3.02	49.55	53.98	52.85
	105	54.29	39.01	29.24	30.07	14.95	5.57	51.99	51.59	50.57
10	89	44.81	32.59	23.07	24.06	11.96	2.78	50.86	51.53	50.56
	79	39.50	27.30	20.82	21.05	8.97	2.78	50.48	49.25	49.81
	105	30.38	16.49	8.02	27.98	14.09	5.62	32.73	33.24	33.04
1	89	24.44	13.32	4.86	22.39	11.27	2.81	32.47	33.24	33.04
	79	21.42	10.28	4.64	19.59	8.45	2.81	31.85	32.72	33.04

Table A2: Performance scores for ViT-B_16 with Random Initialization on CIFAR10 and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		65.45			-			97.96	
100	89		55.87			-			97.10	
	79		49.74			-			97.09	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	152.17	144.45	134.64	16.73	11.36	5.05	97.67	97.76	97.39
90	89	129.03	122.73	113.18	13.39	9.09	2.52	97.67	97.70	97.33
	79	114.02	107.35	100.41	11.71	6.81	2.52	97.52	97.70	97.37
	105	130.57	118.59	120.97	15.93	7.69	3.50	97.90	97.67	97.56
80	89	110.58	100.79	102.00	12.74	6.15	1.75	97.81	97.67	97.30
	79	97.65	88.28	90.38	11.15	4.61	1.75	97.57	97.39	97.17
	105	118.97	113.80	107.29	15.44	7.99	3.26	97.74	97.66	97.69
70	89	100.80	96.79	90.51	12.35	6.39	1.63	97.58	97.60	97.58
	79	89.10	84.82	80.31	10.80	4.79	1.63	97.62	97.50	97.44
	105	102.69	96.62	91.03	15.36	7.91	2.98	97.92	97.80	97.56
60	89	86.85	76.66	76.65	12.29	6.33	1.49	97.92	97.80	97.54
	79	76.71	71.75	67.99	10.75	4.74	1.49	97.82	97.36	97.54
	105	84.95	77.55	75.40	13.88	6.59	2.77	97.82	97.59	97.55
50	89	71.74	65.81	63.35	11.11	5.27	1.38	97.73	97.54	97.46
	79	63.30	57.45	56.14	9.72	3.95	1.38	97.44	97.45	97.27
	105	72.94	64.51	61.54	13.65	6.51	2.70	97.69	97.55	97.59
40	89	61.60	54.78	51.64	10.92	5.21	1.35	97.61	97.55	97.59
	79	54.44	47.81	45.89	9.55	3.90	1.35	97.51	97.31	97.47
	105	55.72	49.19	46.30	13.29	6.42	2.70	97.49	97.52	97.25
30	89	46.86	41.66	38.58	10.63	5.13	1.35	97.46	97.52	97.16
	79	41.35	36.16	34.28	9.30	3.85	1.35	97.16	97.28	97.25
	105	42.39	34.51	31.11	13.92	6.43	2.68	97.28	97.16	97.07
20	89	35.39	29.07	25.56	11.13	5.14	1.34	97.28	97.02	96.98
	79	31.15	24.97	22.71	9.74	3.86	1.34	97.16	96.82	96.91
	105	27.07	20.67	17.29	12.87	6.28	2.67	96.67	96.70	96.38
10	89	22.47	17.36	13.87	10.30	5.03	1.33	96.56	96.70	96.36
	79	19.83	14.73	12.47	9.01	3.77	1.33	96.28	96.67	96.27
	105	14.30	7.87	4.13	12.87	6.43	2.70	92.31	93.00	93.14
1	89	11.52	6.38	2.58	10.30	5.15	1.35	92.31	93.00	93.12
	79	10.10	4.95	2.44	9.01	3.86	1.35	92.31	93.00	93.00

Table A3: Performance scores for ResNet50V2 with ImageNet21k weights Initialization on CIFAR10 and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		65.45			-			88.90	
100	89		55.87			-			88.64	
	79		49.74			-			88.10	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	152.17	144.45	134.64	16.73	11.36	5.05	89.59	89.46	89.64
90	89	129.03	122.73	113.18	13.39	9.09	2.52	89.31	89.24	89.48
	79	114.02	107.35	100.41	11.71	6.81	2.52	89.03	89.10	89.03
	105	130.57	118.59	120.97	15.93	7.69	3.50	88.70	89.06	89.33
80	89	110.58	100.79	102.00	12.74	6.15	1.75	88.37	88.80	89.18
	79	97.65	88.28	90.38	11.15	4.61	1.75	88.05	88.34	88.47
	105	118.97	113.80	107.29	15.44	7.99	3.26	87.31	87.38	87.97
70	89	100.80	96.79	90.51	12.35	6.39	1.63	86.97	86.75	87.91
	79	89.10	84.82	80.31	10.80	4.79	1.63	86.51	86.54	87.04
	105	102.69	96.62	91.03	15.36	7.91	2.98	86.14	86.01	86.79
60	89	86.85	76.66	76.65	12.29	6.33	1.49	85.94	85.47	86.55
	79	76.71	71.75	67.99	10.75	4.74	1.49	85.24	84.98	85.80
	105	84.95	77.55	75.40	13.88	6.59	2.77	83.31	83.70	84.60
50	89	71.74	65.81	63.35	11.11	5.27	1.38	82.79	83.70	84.25
	79	63.30	57.45	56.14	9.72	3.95	1.38	82.14	82.56	83.81
	105	72.94	64.51	61.54	13.65	6.51	2.70	80.91	80.98	81.99
40	89	61.60	54.78	51.64	10.92	5.21	1.35	80.32	80.66	81.76
	79	54.44	47.81	45.89	9.55	3.90	1.35	79.52	79.75	81.21
	105	55.72	49.19	46.30	13.29	6.42	2.70	78.02	77.77	77.76
30	89	46.86	41.66	38.58	10.63	5.13	1.35	77.20	77.25	77.69
	79	41.35	36.16	34.28	9.30	3.85	1.35	76.16	76.46	77.06
	105	42.39	34.51	31.11	13.92	6.43	2.68	73.00	72.28	71.22
20	89	35.39	29.07	25.56	11.13	5.14	1.34	72.50	71.75	71.16
	79	31.15	24.97	22.71	9.74	3.86	1.34	70.47	70.32	70.23
	105	27.07	20.67	17.29	12.87	6.28	2.67	63.40	62.56	61.27
10	89	22.47	17.36	13.87	10.30	5.03	1.33	62.77	62.45	61.27
	79	19.83	14.73	12.47	9.01	3.77	1.33	62.61	61.60	60.73
	105	14.30	7.87	4.13	12.87	6.43	2.70	27.29	27.72	28.98
1	89	11.52	6.38	2.58	10.30	5.15	1.35	27.24	27.72	28.80
	79	10.10	4.95	2.44	9.01	3.86	1.35	27.24	26.84	28.41

Table A4: Performance scores for ResNet50V2 with Random Initialization on CIFAR10 and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
			147.46			-			99.01	
100	89		125.88			-			98.61	
	79		112.06			-			98.86	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	209.07	193.99	183.22	30.87	15.80	5.03	99.08	99.03	99.06
80	89	176.77	164.72	154.59	24.70	12.64	2.51	99.08	99.03	99.06
	79	156.05	143.92	136.95	21.61	9.48	2.51	99.04	99.03	99.06
	105	138.02	123.88	116.13	26.94	12.80	5.04	99.04	98.96	98.94
50	89	116.32	105.01	97.29	21.55	10.24	2.52	99.03	98.96	98.94
	79	102.61	91.43	86.27	18.85	7.68	2.52	99.03	98.96	98.94
	105	48.38	35.97	28.16	25.23	12.82	5.01	98.87	98.72	98.53
10	89	40.02	30.09	22.34	20.19	10.26	2.50	98.87	98.72	98.53
	79	35.29	25.32	20.14	17.66	7.69	2.50	98.87	98.71	98.53
	105	27.15	14.95	7.25	24.84	12.63	4.93	97.90	97.88	97.57
1	89	21.85	12.09	4.45	19.87	10.10	2.46	97.89	97.88	97.57
	79	19.15	9.34	4.23	17.38	8.45	2.46	97.87	97.87	97.57

Table A5: Performance scores for ViT-B_16 with ImageNet 21K weights on CIFAR10 and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		147.46			-			66.84	
100	89		125.88			-			62.21	
	79		112.06			-			64.15	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	209.07	193.99	183.22	30.87	15.80	5.03	66.79	67.05	67.22
80	89	176.77	164.72	154.59	24.70	12.64	2.51	66.33	66.70	66.76
	79	156.05	143.92	136.95	21.61	9.48	2.51	65.92	65.45	66.08
	105	138.02	123.88	116.13	26.94	12.80	5.04	62.90	63.56	63.53
50	89	116.32	105.01	97.29	21.55	10.24	2.52	62.66	63.00	63.08
	79	102.61	91.43	86.27	18.85	7.68	2.52	61.57	62.74	62.49
	105	48.38	35.97	28.16	25.23	12.82	5.01	50.53	50.16	49.04
10	89	40.02	30.09	22.34	20.19	10.26	2.50	50.45	50.00	48.95
	79	35.29	25.32	20.14	17.66	7.69	2.50	50.06	49.44	48.45
	105	27.15	14.95	7.25	24.84	12.63	4.93	27.77	27.55	30.00
1	89	21.85	12.09	4.45	19.87	10.10	2.46	27.51	27.55	30.00
	79	19.15	9.34	4.23	17.38	8.45	2.46	22.27	27.55	30.00

Table A6: Performance scores for ViT-B_16 with random weights weights on CIFAR10 and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		65.45			-			97.96	
100	89		55.87			-			97.10	
	79		49.74			-			97.09	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	124.89	119.48	115.79	11.43	6.02	2.33	97.93	98.00	97.88
80	89	105.98	101.65	97.99	9.15	4.82	1.16	97.93	97.94	97.81
	79	93.61	89.21	86.77	8.00	3.61	1.16	97.90	97.93	97.80
	105	82.37	77.27	73.27	11.64	6.54	2.54	97.87	97.90	97.63
50	89	69.65	65.57	61.61	9.31	5.23	1.27	97.85	97.90	97.58
	79	61.47	57.25	54.59	8.15	3.92	1.27	97.80	97.72	97.58
	105	30.69	22.63	18.14	15.95	7.9	3.40	97.70	97.30	96.65
10	89	25.39	18.94	14.33	12.76	6.31	1.70	97.60	97.25	96.65
	79	22.39	15.96	12.92	5.15	4.73	1.70	97.48	97.07	96.56
	105	11.31	6.65	3.62	9.84	5.17	2.15	93.14	92.31	92.91
1	89	9.13	5.40	2.33	7.87	4.14	1.07	92.93	92.29	92.91
	79	8.01	4.23	2.19	6.89	3.10	1.07	92.93	92.29	92.91

Table A7: Performance scores for ResNet50V2 with ImageNet21k weights Initialization on CIFAR10 and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		65.45			-			88.90	
100	89		55.87			-			88.64	
	79		49.74			-			88.10	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	124.89	119.48	115.79	11.43	6.02	2.33	89.55	89.78	89.55
80	89	105.98	101.65	97.99	9.15	4.82	1.16	89.40	89.69	89.52
	79	93.61	89.21	86.77	8.00	3.61	1.16	89.01	89.31	89.40
	105	82.37	77.27	73.27	11.64	6.54	2.54	86.81	86.26	85.09
50	89	69.65	65.57	61.61	9.31	5.23	1.27	86.65	86.19	85.09
	79	61.47	57.25	54.59	8.15	3.92	1.27	86.08	85.74	84.86
	105	30.69	23.33	18.57	15.95	8.59	3.83	63.91	64.35	60.61
10	89	25.39	19.50	14.54	12.76	6.87	1.91	63.59	61.49	60.61
	79	22.39	16.38	13.14	5.15	3.77	1.91	61.99	61.49	60.61
	105	11.31	6.65	3.62	9.84	5.17	2.15	23.74	24.92	27.15
1	89	9.13	5.40	2.33	7.87	4.14	1.07	23.74	24.92	27.08
	79	8.01	4.23	2.19	6.89	3.10	1.07	22.83	24.92	26.76

Table A8: Performance scores for ResNet50V2 with Random Initialization on CIFAR10 and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
			311.64			-			89.44	
100	105									
	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	575.97	474.68	421.88	186.06	84.77	31.97	89.46	89.28	89.23
80	89	480.69	399.65	347.82	148.85	67.81	15.98	89.46	89.28	89.23
	79	425.21	345.83	310.95	130.24	50.86	15.98	89.41	89.28	89.12
	105	429.41	328.12	275.32	186.06	84.77	31.97	89.97	89.59	89.18
50	89	356.91	275.87	224.04	148.85	67.81	15.98	89.90	89.59	89.16
	79	314.60	235.21	200.34	130.24	50.86	15.98	89.90	89.42	89.04
	105	332.62	231.33	178.53	186.06	84.77	31.97	88.48	89.05	89.01
30	89	272.63	191.59	139.76	148.85	67.81	15.98	88.40	88.81	88.83
	79	240.85	161.47	126.60	130.24	50.86	15.98	88.33	88.77	88.70
	105	235.84	134.54	81.75	186.06	84.77	31.97	84.97	84.46	86.00
10	89	190.99	109.95	58.12	148.85	67.81	15.98	84.62	84.46	86.00
	79	167.11	87.73	52.85	130.24	50.86	15.98	83.99	83.95	86.00
	105	191.59	90.30	37.50	186.06	84.77	31.97	48.30	39.39	33.73
1	89	153.59	72.55	20.72	148.85	67.81	15.98	48.04	38.83	33.70
	79	134.45	55.07	20.20	130.24	50.86	15.98	44.97	38.00	33.70

Table A9: Performance scores for ViT-B_16 with ImageNet-21k weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
			311.64			-			89.44	
100	105									
	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	518.21	490.10	452.71	120.78	61.71	24.32	89.50	90.19	89.96
80	89	434.47	415.78	378.57	96.62	49.37	12.16	89.50	90.19	89.96
	79	384.77	361.86	336.99	84.55	37.03	12.16	89.50	90.19	89.96
	105	371.64	331.07	292.52	128.95	63.67	25.12	89.82	90.02	89.65
50	89	310.69	279.62	241.24	103.16	50.93	12.56	89.82	90.02	89.65
	79	274.16	240.89	215.25	90.26	38.20	12.56	89.68	90.02	89.65
	105	274.86	220.17	185.96	120.29	59.18	24.97	89.89	89.72	89.22
30	89	226.41	185.07	150.21	96.23	47.34	12.48	89.89	89.72	89.22
	79	200.42	157.64	134.62	84.20	35.51	12.48	89.76	89.72	89.22
	105	178.07	113.64	78.59	119.51	59.07	24.02	89.07	87.98	88.06
10	89	144.77	94.03	58.78	95.61	47.26	12.01	88.88	87.98	87.51
	79	126.67	77.02	53.58	83.66	35.44	12.01	88.77	87.98	87.51
	105	133.82	65.22	29.64	128.29	59.76	24.19	59.51	54.25	49.46
1	89	107.37	52.49	16.77	102.63	47.81	12.09	59.09	50.07	41.88
	79	94.02	40.01	16.25	89.80	35.86	12.09	58.49	43.18	38.26

Table A10: Performance scores for ViT-B_16 with ImageNet-21k weights on Tiny Imagenet and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		311.64			-			89.44	
100	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	472.06	490.10	452.71	82.14	39.50	15.48	89.60	90.19	89.96
80	89	397.56	415.78	378.57	65.71	31.60	7.74	89.56	90.19	89.96
	79	352.47	361.86	336.99	57.50	23.70	7.74	89.56	90.19	89.96
	105	325.49	331.07	292.52	82.14	39.50	15.48	89.69	90.02	89.65
50	89	273.77	279.62	241.24	65.71	31.60	7.74	89.56	90.02	89.65
	79	241.86	240.89	215.25	57.50	23.70	7.74	89.49	90.02	89.65
	105	228.71	220.17	185.96	82.14	39.50	15.48	89.82	89.72	89.22
30	89	189.50	185.07	150.21	65.71	31.60	7.74	89.77	89.72	89.22
	79	168.11	157.64	134.62	57.50	23.70	7.74	89.73	89.72	89.22
	105	131.92	95.83	71.01	82.14	39.50	15.48	88.89	88.20	87.75
10	89	107.85	79.78	54.99	65.71	31.60	7.74	88.81	88.20	87.38
	79	94.37	66.33	49.79	57.50	23.70	7.74	88.63	88.20	87.38
	105	87.67	45.03	21.01	82.14	39.50	15.48	45.59	-	39.03
1	89	70.45	36.34	12.48	65.71	31.60	7.74	45.43	-	39.03
	79	61.71	27.91	11.95	57.50	23.70	7.74	44.75	-	39.03

Table A11: Performance scores for ViT-B_16 with ImageNet-21k weights on Tiny Imagenet and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		311.64			-			89.44	
100	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	389.91	389.91	389.91	0.00	0.00	0.00	89.33	90.04	89.93
80	89	331.84	331.84	331.84	0.00	0.00	0.00	89.33	90.04	89.93
	79	294.97	294.97	294.97	0.00	0.00	0.00	89.33	90.04	89.93
	105	243.35	243.35	243.35	0.00	0.00	0.00	88.67	89.81	89.60
50	89	208.05	208.05	208.05	0.00	0.00	0.00	88.67	89.81	89.60
	79	184.35	184.35	184.35	0.00	0.00	0.00	88.67	89.81	89.60
	105	146.56	146.56	146.56	0.00	0.00	0.00	88.44	89.24	89.39
30	89	123.78	123.78	123.78	0.00	0.00	0.00	88.44	89.24	89.39
	79	110.61	110.61	110.61	0.00	0.00	0.00	88.44	89.24	89.39
	105	49.77	49.77	49.77	0.00	0.00	0.00	86.69	87.62	87.69
10	89	42.13	42.13	42.13	0.00	0.00	0.00	86.69	87.62	87.69
	79	36.87	36.87	36.87	0.00	0.00	0.00	86.69	87.62	87.69
	105	5.53	5.53	5.53	0.00	0.00	0.00	40.24	41.72	41.87
1	89	4.74	4.74	4.74	0.00	0.00	0.00	39.96	40.57	39.95
	79	4.21	4.21	4.21	0.00	0.00	0.00	39.41	40.21	39.84

Table A12: Performance scores for ViT-B_16 with ImageNet-21k weights on Tiny Imagenet and its Effective data per epoch (EDPE), method=Random

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		111.61			-			79.02	
100	89		95.45			-			77.94	
	79		84.60			-			77.54	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	389.09	326.49	279.40	136.72	74.11	27.02	79.76	79.70	79.16
80	89	324.16	274.08	228.30	109.37	59.29	13.51	79.52	79.70	79.16
	79	286.62	235.39	204.43	95.70	44.47	13.51	79.24	79.29	79.16
	105	294.23	231.62	184.53	136.72	74.11	27.02	80.87	80.23	78.89
50	89	244.04	193.96	148.18	109.37	59.29	13.51	80.71	79.94	78.77
	79	215.03	163.79	132.84	95.70	44.47	13.51	80.41	79.66	78.64
	105	231.58	168.98	121.89	136.72	74.11	27.02	78.94	79.19	77.71
30	89	189.49	139.41	93.63	109.37	59.29	13.51	78.89	78.81	77.68
	79	167.30	116.06	85.11	95.70	44.47	13.51	78.55	78.18	77.28
	105	168.94	106.33	59.24	136.72	74.11	27.02	72.21	73.59	73.18
10	89	136.65	86.56	40.78	109.37	59.29	13.51	71.71	73.59	73.12
	79	119.57	68.33	37.37	95.70	44.47	13.51	71.71	73.59	72.89
	105	140.30	77.69	30.60	136.72	74.11	27.02	60.83	55.63	49.00
1	89	112.44	62.36	16.58	109.37	59.29	13.51	59.86	54.56	49.00
	79	98.43	47.19	16.24	95.70	44.47	13.51	58.10	50.75	49.00

Table A13: Performance scores for ResNet50V2 with ImageNet-21k Initialization on Tiny-ImageNet and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		111.61			-			79.02	
100	89		95.45			-			77.94	
	79		84.60			-			77.54	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	321.22	283.43	258.94	77.97	40.18	15.70	79.40	79.42	79.16
80	89	269.39	239.16	214.86	62.38	32.15	7.85	79.37	79.38	79.16
	79	238.59	208.12	191.86	0.00	0.00	0.00	79.37	79.16	79.16
	105	229.79	192.00	167.51	77.97	40.18	15.70	80.41	80.04	78.61
50	89	192.17	161.94	137.64	62.38	32.15	7.85	80.28	79.84	78.61
	79	169.59	139.12	122.86	54.58	24.11	7.85	79.64	79.01	78.61
	105	169.40	131.62	107.13	77.97	40.18	15.70	81.09	80.29	77.74
30	89	139.60	109.37	85.07	62.38	32.15	7.85	80.96	80.12	77.65
	79	123.58	93.11	76.85	54.58	24.11	7.85	80.46	79.05	77.54
	105	109.02	71.24	46.75	77.97	40.18	15.70	80.24	78.43	75.36
10	89	88.66	58.43	34.13	62.38	32.15	7.85	79.98	78.20	75.36
	79	77.58	47.11	30.85	54.58	24.11	7.85	79.11	76.95	74.98
	105	81.42	43.64	19.15	77.97	40.18	15.70	72.49	68.88	60.88
1	89	65.33	35.10	10.80	62.38	32.15	7.85	71.71	68.61	60.80
	79	57.21	26.74	10.47	54.58	24.11	7.85	70.50	66.61	60.06

Table A14: Performance scores for ResNet50V2 with ImageNet-21k weights on Tiny Imagenet and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		111.61			-			79.02	
100	89		95.45			-			77.94	
	79		84.60			-			77.54	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	309.69	278.64	259.50	62.48	31.44	12.29	79.09	79.09	79.09
80	89	261.88	237.04	218.03	49.99	25.15	6.14	79.09	79.09	79.09
	79	232.08	207.21	194.49	43.74	18.86	6.14	79.09	79.09	79.09
	105	217.87	186.82	167.68	62.48	31.44	12.29	78.11	78.11	78.11
50	89	181.16	156.32	137.31	49.99	25.15	6.14	78.11	78.11	78.11
	79	161.45	136.58	123.86	43.74	18.86	6.14	78.11	78.11	78.11
	105	154.30	123.26	104.11	62.48	31.44	12.29	76.73	76.11	75.55
30	89	127.34	102.51	83.50	49.99	25.15	6.14	76.42	75.81	75.55
	79	111.00	86.13	73.41	43.74	18.86	6.14	76.42	75.74	75.55
	105	94.27	63.225	44.07	62.48	31.44	12.29	75.08	72.32	71.22
10	89	76.89	52.06	33.05	49.99	25.15	6.14	74.33	72.32	71.22
	79	67.28	42.40	29.69	43.74	18.86	6.14	74.33	72.32	71.22
	105	66.01	34.97	15.82	62.48	31.44	12.29	63.99	64.10	57.57
1	89	53.01	28.18	9.17	49.99	25.15	6.14	63.71	62.11	57.57
	79	46.43	21.55	8.83	43.74	18.86	6.14	63.71	61.24	57.57

Table A15: Performance scores for ResNet50V2 with ImageNet-21k weights on Tiny Imagenet and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		111.61			-			79.02	
100	89		95.45			-			77.94	
	79		84.60			-			77.54	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	250.73	250.73	250.73	0.00	0.00	0.00	78.81	78.81	78.81
80	89	211.88	211.88	211.88	0.00	0.00	0.00	78.81	78.81	78.81
	79	188.34	188.34	188.34	0.00	0.00	0.00	78.81	78.81	78.81
	105	155.38	155.38	155.38	0.00	0.00	0.00	78.17	78.17	78.17
50	89	131.17	131.17	131.17	0.00	0.00	0.00	78.17	78.17	78.17
	79	117.71	117.71	117.71	0.00	0.00	0.00	78.17	78.17	78.17
	105	95.35	95.35	95.35	0.00	0.00	0.00	75.59	75.59	75.59
30	89	80.72	80.72	80.72	0.00	0.00	0.00	75.59	75.59	75.59
	79	70.62	70.62	70.62	0.00	0.00	0.00	75.59	75.59	75.59
	105	31.78	31.78	31.78	0.00	0.00	0.00	72.07	71.52	71.51
10	89	26.90	26.90	26.90	0.00	0.00	0.00	72.07	71.52	71.51
	79	23.54	23.54	23.54	0.00	0.00	0.00	72.07	71.52	71.51
	105	3.53	3.53	3.53	0.00	0.00	0.00	52.21	57.02	56.76
1	89	3.02	3.02	3.02	0.00	0.00	0.00	52.21	57.02	56.76
	79	2.69	2.69	2.69	0.00	0.00	0.00	52.21	57.02	56.76

Table A16: Performance scores for ResNet50V2 with ImageNet-21k weights on Tiny Imagenet and its Effective data per epoch (EDPE), method=Random

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		113.96			-			74.83	
100	89		97.68			-			74.79	
	79		86.82			-			74.56	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	398.84	307.91	265.86	161.91	70.98	28.93	75.76	75.24	74.88
80	89	331.17	258.42	216.10	129.53	56.78	14.46	75.51	75.17	74.88
	79	292.57	221.82	193.70	113.33	42.59	14.46	75.24	75.05	74.47
	105	309.78	218.85	176.80	161.91	70.98	28.93	75.35	74.73	73.99
50	89	255.95	183.21	140.89	129.53	56.78	14.46	75.11	74.36	73.93
	79	225.36	154.61	126.49	113.33	42.59	14.46	75.01	74.16	73.75
	105	250.97	160.04	117.99	161.91	70.98	28.93	72.10	72.71	71.72
30	89	204.74	132.00	89.68	129.53	56.78	14.46	72.04	72.29	71.55
	79	180.55	109.80	81.68	113.33	42.59	14.46	71.46	71.90	71.02
	105	192.15	101.23	59.18	161.91	70.98	28.93	59.15	60.22	62.17
10	89	155.13	82.39	40.07	129.53	56.78	14.46	57.16	60.22	62.03
	79	135.74	64.99	36.87	113.33	42.59	14.46	57.16	60.22	60.43
	105	165.27	74.34	32.29	161.91	70.98	28.93	17.15	18.36	19.08
1	89	132.41	59.66	17.34	129.53	56.78	14.46	16.83	17.88	18.98
	79	115.89	45.15	17.02	113.33	42.59	14.46	14.95	17.38	18.98

Table A17: Performance scores for MobileNetV3 with ImageNet-1k weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		113.96			-			74.83	
100	89		97.68			-			74.79	
	79		86.82			-			74.56	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	314.47	275.92	252.20	77.55	38.99	15.27	75.39	75.58	74.92
80	89	263.68	232.83	209.27	62.04	31.19	7.63	75.31	75.29	74.91
	79	233.52	202.63	186.87	54.28	23.39	7.63	75.25	74.87	74.82
	105	225.42	186.86	163.14	77.55	38.99	15.27	75.75	75.23	73.66
50	89	188.46	157.62	134.06	62.04	31.19	7.63	75.64	75.15	73.66
	79	166.30	135.41	119.66	54.28	23.39	7.63	75.39	74.87	73.66
	105	166.60	128.05	104.33	77.55	38.99	15.27	75.47	74.60	71.82
30	89	137.25	106.40	82.85	62.04	31.19	7.63	75.40	74.44	71.71
	79	121.49	90.60	74.85	54.28	23.39	7.63	74.98	73.63	71.52
	105	107.79	69.23	45.52	77.55	38.99	15.27	72.42	70.02	66.05
10	89	87.64	56.79	33.24	62.04	31.19	7.63	71.61	69.70	65.85
	79	76.68	45.79	30.04	54.28	23.39	7.63	70.60	68.96	65.54
	105	80.91	42.35	18.63	77.55	38.99	15.27	29.35	29.78	27.18
1	89	64.92	34.07	10.51	62.04	31.19	7.63	27.61	29.47	26.21
	79	56.84	25.95	10.19	54.28	23.39	7.63	25.58	27.96	25.01

Table A18: Performance scores for MobileNetV3 with ImageNet-1k weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		113.96			-			74.83	
100	89		97.68			-			74.79	
	79		86.82			-			74.56	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	275.87	256.01	244.63	38.94	19.08	7.70	75.87	75.63	75.22
80	89	232.79	216.90	205.49	31.15	15.26	3.85	75.71	75.48	75.04
	79	206.49	190.68	183.08	27.25	11.44	3.85	75.45	75.24	74.76
	105	186.81	166.95	155.57	38.94	19.08	7.70	75.45	75.36	73.88
50	89	157.57	130.27	130.27	31.15	15.26	3.85	74.95	74.62	73.31
	79	139.28	115.87	126.49	27.25	11.44	3.85	74.60	74.02	73.31
	105	127.99	108.14	96.76	38.94	19.08	7.70	75.15	73.94	72.12
30	89	106.36	90.48	79.06	31.15	15.26	3.85	73.49	72.62	71.42
	79	94.47	78.66	71.06	27.25	11.44	3.85	72.91	71.40	71.42
	105	69.18	49.32	37.94	38.94	19.08	7.70	68.37	65.32	63.78
10	89	56.75	40.87	29.45	31.15	15.26	3.85	67.78	64.93	63.68
	79	49.66	33.85	26.25	27.25	11.44	3.85	66.93	64.93	63.19
	105	42.30	22.44	11.06	38.94	19.08	7.70	20.68	26.97	27.05
1	89	34.03	18.14	6.73	31.15	15.26	3.85	17.57	25.78	26.42
	79	29.81	14.00	6.41	27.25	11.44	3.85	14.95	23.64	25.71

Table A19: Performance scores for MobileNetV3 with ImageNet-1k weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		113.96			-			74.83	
100	89		97.68			-			74.79	
	79		86.82			-			74.56	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	236.92	236.92	236.92	0.0	0.0	0.0	74.44	74.12	74.92
80	89	201.64	201.64	201.64	0.0	0.0	0.0	74.40	74.10	74.91
	79	179.23	179.23	179.23	0.0	0.0	0.0	74.12	74.00	74.82
	105	147.87	147.87	147.87	0.0	0.0	0.0	72.34	72.39	73.85
50	89	126.42	126.42	126.42	0.0	0.0	0.0	72.28	72.39	73.72
	79	112.02	112.02	112.02	0.0	0.0	0.0	72.05	72.39	73.69
	105	89.05	89.05	89.05	0.0	0.0	0.0	70.60	70.68	72.35
30	89	75.21	75.21	75.21	0.0	0.0	0.0	70.52	70.60	72.26
	79	67.21	67.21	67.21	0.0	0.0	0.0	69.91	70.33	72.22
	105	30.24	30.24	30.24	0.0	0.0	0.0	62.58	64.29	65.77
10	89	25.60	25.60	25.60	0.0	0.0	0.0	62.15	64.29	65.77
	79	22.40	22.40	22.40	0.0	0.0	0.0	61.90	63.87	65.77
	105	3.36	3.36	3.36	0.0	0.0	0.0	22.72	25.81	29.05
1	89	2.88	2.88	2.88	0.0	0.0	0.0	22.19	25.35	28.30
	79	2.56	2.56	2.56	0.0	0.0	0.0	21.40	24.47	27.17

Table A20: Performance scores for MobileNetV3 with ImageNet-1k weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=Random

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
			311.64			-			89.44	
100	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	1150.82	307.91	265.86	226.14	103.75	40.27	90.89	-	-
80	89	967.87	258.42	216.10	180.91	83.00	20.13	90.78	-	-
	79	857.81	221.82	193.70	158.30	62.25	20.13	90.60	-	-
	105	803.24	680.85	617.37	226.14	103.75	40.27	91.51	91.49	90.90
50	89	674.32	576.41	513.54	180.91	83.00	20.13	91.51	91.41	90.78
	79	595.50	499.44	457.33	158.30	62.25	20.13	91.28	91.14	90.62
	105	573.72	451.32	387.84	226.14	103.75	40.27	91.28	91.05	90.42
30	89	474.46	376.55	313.68	180.91	83.00	20.13	91.15	90.70	90.33
	79	420.62	324.57	282.45	158.30	62.25	20.13	90.65	90.56	90.24
	105	343.56	221.79	155.19	226.14	103.75	40.27	86.60	86.83	88.26
10	89	280.85	182.93	113.82	180.91	83.00	20.13	86.40	86.83	87.60
	79	245.74	149.69	107.57	158.30	62.25	20.13	85.43	86.83	87.46
	105	239.26	116.87	53.38	226.14	103.75	40.27	76.14	72.53	66.49
1	89	192.16	94.24	31.37	180.91	83.00	20.13	75.39	72.00	66.15
	79	168.29	72.24	30.12	158.30	62.25	20.13	72.97	69.72	65.89

Table A21: Performance scores for Swin TransformerV2 with ImageNet-21K and ImageNet1k fine-tuned weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		311.64			-			89.44	
100	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	398.84	307.91	265.86	125.24	62.45	25.20	-	-	-
80	89	331.17	258.42	216.10	100.19	49.96	12.60	-	-	-
	79	292.57	221.82	193.70	87.67	37.47	20.13	-	-	-
	105	702.34	639.55	602.30	125.24	62.45	25.20	91.44	91.43	90.76
50	89	593.60	543.37	506.01	100.19	49.96	12.60	91.34	91.34	90.69
	79	524.87	474.66	449.80	87.67	37.47	20.13	91.24	91.21	90.69
	105	472.82	410.02	372.78	125.24	62.45	25.20	91.90	91.48	90.66
30	89	393.74	343.50	306.15	100.19	49.96	12.60	91.71	91.43	90.58
	79	349.99	299.79	274.92	87.67	37.47	20.13	91.46	91.21	90.56
	105	243.29	180.49	143.25	125.24	62.45	25.20	91.48	90.93	89.56
10	89	200.13	149.89	112.53	100.19	49.96	12.60	91.31	90.66	89.56
	79	175.11	124.91	100.04	87.67	37.47	20.13	91.17	90.62	89.51
	105	138.36	75.56	38.32	125.24	62.45	25.20	85.37	83.72	77.42
1	89	111.44	61.20	23.84	100.19	49.96	12.60	84.96	83.59	77.42
	79	97.66	47.46	22.59	87.67	37.47	20.13	84.00	81.87	76.37

Table A22: Performance scores for Swin TransformerV2 with ImageNet-21K and ImageNet1k fine-tuned weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
100	105		311.64			-			89.44	
	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
80	105	1024.94	307.91	265.86	226.14	103.75	40.27	90.97	-	-
	89	867.17	258.42	216.10	180.91	83.00	20.13	90.95	-	-
	79	769.70	221.82	193.70	158.30	62.25	20.13	90.95	-	-
50	105	677.37	680.85	617.37	226.14	103.75	40.27	91.47	91.49	90.90
	89	573.62	576.41	513.54	180.91	83.00	20.13	91.47	91.41	90.78
	79	507.39	499.44	457.33	158.30	62.25	20.13	91.27	91.14	90.62
30	105	447.84	451.32	387.84	226.14	103.75	40.27	91.87	91.05	90.42
	89	373.76	376.55	313.68	180.91	83.00	20.13	91.69	90.70	90.33
	79	332.51	324.57	282.45	158.30	62.25	20.13	91.69	90.56	90.24
10	105	218.32	220.54	155.19	226.14	103.75	40.27	91.63	86.83	88.26
	89	180.15	182.93	113.82	180.91	83.00	20.13	91.05	86.83	87.60
	79	157.63	149.69	107.57	158.30	62.25	20.13	90.81	86.83	87.46
1	105	113.39	116.87	53.38	100.27	103.75	40.27	86.41	72.53	66.49
	89	91.46	94.24	31.37	80.22	83.00	20.13	85.82	72.00	66.15
	79	80.18	72.24	30.12	70.19	62.25	20.13	84.47	69.72	65.89

Table A23: Performance scores for Swin TransformerV2 with ImageNet-21K and ImageNet1k fine-tuned weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
100	105		311.64			-			89.44	
	89		266.52			-			88.94	
	79		236.23			-			88.94	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
80	105	924.67	924.67	924.67	0.0	0.0	0.0	-	-	-
	89	786.95	786.95	786.95	0.0	0.0	0.0	-	-	-
	79	699.51	699.51	699.51	0.0	0.0	0.0	-	-	-
50	105	577.09	577.09	577.09	0.0	0.0	0.0	90.00	91.43	90.76
	89	493.40	493.40	493.40	0.0	0.0	0.0	90.00	91.34	90.69
	79	437.19	437.19	437.19	0.0	0.0	0.0	90.00	91.21	90.69
30	105	347.57	347.57	347.57	0.0	0.0	0.0	90.02	91.48	90.66
	89	293.54	343.50	306.15	0.0	0.0	0.0	90.02	91.43	90.58
	79	262.31	299.79	274.92	0.0	0.0	0.0	90.02	91.21	90.56
10	105	118.04	118.04	118.04	0.0	0.0	0.0	87.77	90.93	89.56
	89	99.93	149.89	112.53	0.0	0.0	0.0	87.77	90.66	89.56
	79	87.43	124.91	100.04	0.0	0.0	0.0	87.77	90.62	89.51
1	105	13.11	13.11	13.11	0.0	0.0	0.0	64.97	83.72	77.42
	89	11.24	11.24	11.24	0.0	0.0	0.0	64.91	83.59	77.42
	79	9.99	9.99	9.99	0.0	0.0	0.0	64.52	81.87	76.37

Table A24: Performance scores for Swin TransformerV2 with ImageNet-21K and ImageNet1k fine-tuned weights Initialization on Tiny Imagenet and its Effective data per epoch (EDPE), method=Random

A.4   CRAIG, GradMatch, GLISTER and Random Results on Oxford-IIIT Pet

To further study the effect on ImageNet21k pretaining and random initialized models with the coreset methods. We further conducted experiments on the Oxford-IIIT Pet dataset. Each experiment in the tables are conducted five times and results reported in the table are average with their standard deviation. Tables [A25, A26] shows comparative study of these models, and we used Fig. 6 to show that when transformers are initialized with ImageNet21k they outperform CNNs across all the EDPE values and coreset methods. When the models are randomly initialized, it can be seen that transformers is suffering and CNNs are outperforming them.

CRAIG, GradMatch, GLISTER and Random Results on UltraMNIST To examine how well transformers can learn a robust representation in a dataset, we used UltraMNIST to compare ViT_B16 and ResNet50. Tables [A27, A28, A29, A30, A31, A32, A33, A34] delivers a comparative study for these models, and we used Fig. 8 to exhibit that CNN outperforms across all coreset sizes and that due to nonavailability of similar pretraining on such non-natural image dataset, transformers tend to suffer in learning desired semantic context as mentioned in the section 'Learning semantic coherency in non-natural images' in the main paper. For UltraMNIST, GradMatch seems to perform best when compared to other methods, and random selection performs worst. These comments are uniform with varying SSI values, as shown in Fig. 9 in the main document.

CRAIG, GradMatch, GLISTER and Random Results on Medical Dataset We further expanded our experiments vision to a medical dataset, to study the effect of pretraining when using ViT_B16 and ResNet50 with different coreset sizes. From the tables [A35, A36, A37, A38, A39, A40, A41, A42] and the generated Fig. 7 from these tables assisted us to show that without pretraining CNNs outperform on Transformers for different coreset values using random selection. It was also intriguing to see that random selection is better or almost at par with GradMatch, which is itself superior to other coreset methods.

Model	Baseline	EDPE	Total Training Time	Accuracy (Pretrain)	Accuracy (Random)
ResNet50	Full data	100	5.74	91.04 ± 0.29	21.39 ± 0.98
MobileNetV3	Full data	100	4.68	86.51 ± 0.31	19.94 ± 1.25
Model	Coreset Method	EDPE	Total Training Time	Accuracy (Pretrain)	Accuracy (Random)
ResNet50	CRAIG	1	4.64	44.68 ± 4.17	3.40 ± 0.17
ResNet50	CRAIG	10	4.68	81.76 ± 0.94	3.05 ± 0.45
ResNet50	CRAIG	30	6.56	86.74 ± 0.55	5.08 ± 0.33
ResNet50	CRAIG	50	9.45	88.96 ± 0.33	12.04 ± 1.02
ResNet50	CRAIG	80	12.65	89.45 ± 0.64	19.88 ± 1.04
ResNet50	GradMatch	1	2.37	44.68 ± 4.17	3.24 ± 0.24
ResNet50	GradMatch	10	2.40	84.95 ± 1.34	2.96 ± 0.46
ResNet50	GradMatch	30	4.28	87.92 ± 0.58	5.05 ± 0.64
ResNet50	GradMatch	50	7.17	89.32 ± 0.27	11.91 ± 0.93
ResNet50	GradMatch	80	10.38	89.76 ± 0.28	19.77 ± 0.79
ResNet50	GLISTER	1	1.15	46.87 ± 2.77	3.57 ± 0.31
ResNet50	GLISTER	10	1.18	84.68 ± 0.34	2.81 ± 0.09
ResNet50	GLISTER	30	3.06	88.48 ± 0.48	5.15 ± 0.50
ResNet50	GLISTER	50	5.95	89.28 ± 0.50	13.08 ± 1.21
ResNet50	GLISTER	80	9.16	89.68 ± 0.42	20.16 ± 0.98
ResNet50	Random	1	0.01	44.68 ± 4.17	3.16 ± 0.13
ResNet50	Random	10	0.04	83.83 ± 1.79	3.03 ± 0.22
ResNet50	Random	30	1.92	87.78 ± 0.33	5.55 ± 0.36
ResNet50	Random	50	4.81	89.48 ± 0.49	12.72 ± 1.21
ResNet50	Random	80	8.02	89.66 ± 0.19	20.19 ± 0.81
MobileNetV3	CRAIG	1	3.59	28.53 ± 2.34	2.72 ± 0.00
MobileNetV3	CRAIG	10	3.62	73.06 ± 1.71	2.72 ± 0.00
MobileNetV3	CRAIG	30	5.44	79.41 ± 0.80	2.72 ± 0.00
MobileNetV3	CRAIG	50	9.27	84.13 ± 0.71	2.72 ± 0.00
MobileNetV3	CRAIG	80	13.06	85.82 ± 0.35	8.78 ± 0.90
MobileNetV3	GradMatch	1	1.32	27.42 ± 0.93	2.72 ± 0.00
MobileNetV3	GradMatch	10	1.34	76.94 ± 1.45	2.72 ± 0.00
MobileNetV3	GradMatch	30	3.17	81.96 ± 0.48	2.71 ± 0.01
MobileNetV3	GradMatch	50	6.99	84.76 ± 0.56	2.72 ± 0.00
MobileNetV3	GradMatch	80	10.78	86.22 ± 0.33	8.04 ± 1.53
MobileNetV3	GLISTER	1	4.27	30.85 ± 0.89	2.72 ± 0.00
MobileNetV3	GLISTER	10	4.29	74.53 ± 2.56	2.70 ± 0.03
MobileNetV3	GLISTER	30	6.12	81.23 ± 0.94	2.72 ± 0.00
MobileNetV3	GLISTER	50	9.95	84.67 ± 0.51	2.74 ± 0.05
MobileNetV3	GLISTER	80	13.74	85.96 ± 0.47	6.79 ± 1.18
MobileNetV3	Random	1	0.01	27.82 ± 1.65	2.72 ± 0.00
MobileNetV3	Random	10	0.03	76.92 ± 1.83	2.72 ± 0.00
MobileNetV3	Random	30	1.86	81.89 ± 0.41	2.71 ± 0.01
MobileNetV3	Random	50	5.68	84.57 ± 0.63	2.86 ± 0.26
MobileNetV3	Random	80	9.47	86.06 ± 0.18	7.41 ± 1.36

Table A25: Performance scores for ResNet50 and MobileNetV3 with ImageNet-21k weights and Random initialized on Oxford-IIIT Pet.

Model	Baseline	EDPE	Total Training Time	Accuracy (Pretrain)	Accuracy (Random)
ViT_B16	Full data	100	13.14	93.44 ± 0.11	15.65 ± 0.36
Swin	Full data	100	20.85	95.55 ± 0.28	9.95 ± 0.56

Model	Coreset Method	EDPE	Total Training Time	Accuracy (Pretrain)	Accuracy (Random)
ViT_B16	CRAIG	1	5.42	50.15 ± 6.12	4.36 ± 0.36
ViT_B16	CRAIG	10	5.48	83.36 ± 1.71	5.24 ± 0.43
ViT_B16	CRAIG	30	8.40	91.57 ± 0.11	8.40 ± 0.61
ViT_B16	CRAIG	50	12.23	92.82 ± 0.20	11.86 ± 0.29
ViT_B16	CRAIG	80	16.79	93.15 ± 0.25	14.64 ± 0.43
ViT_B16	GradMatch	1	3.49	48.91 ± 6.23	4.31 ± 0.32
ViT_B16	GradMatch	10	3.55	88.44 ± 1.41	5.93 ± 0.97
ViT_B16	GradMatch	30	6.47	92.17 ± 0.38	8.14 ± 0.35
ViT_B16	GradMatch	50	10.30	92.84 ± 0.37	12.64 ± 0.72
ViT_B16	GradMatch	80	14.86	93.24 ± 0.09	15.18 ± 0.20
ViT_B16	GLISTER	1	1.82	50.33 ± 5.20	4.46 ± 0.42
ViT_B16	GLISTER	10	1.88	86.88 ± 1.40	6.31 ± 0.28
ViT_B16	GLISTER	30	4.80	92.16 ± 0.15	8.73 ± 0.24
ViT_B16	GLISTER	50	8.63	92.81 ± 0.23	12.79 ± 0.61
ViT_B16	GLISTER	80	13.19	93.19 ± 0.09	15.17 ± 0.52
ViT_B16	Random	1	0.02	49.55 ± 5.10	4.44 ± 0.50
ViT_B16	Random	10	0.07	86.84 ± 1.43	6.19 ± 0.35
ViT_B16	Random	30	2.99	91.95 ± 0.30	8.13 ± 0.28
ViT_B16	Random	50	6.83	92.79 ± 0.28	12.26 ± 0.33
ViT_B16	Random	80	11.38	93.05 ± 0.35	14.83 ± 0.15
Swin	CRAIG	1	8.87	55.15 ± 3.99	3.93 ± 0.38
Swin	CRAIG	10	9.12	91.23 ± 0.64	3.92 ± 0.43
Swin	CRAIG	30	15.01	94.16 ± 0.37	5.06 ± 0.20
Swin	CRAIG	50	21.30	94.61 ± 0.34	6.81 ± 0.74
Swin	CRAIG	80	44.01	94.97 ± 0.47	8.99 ± 0.60
Swin	GradMatch	1	5.81	51.93 ± 6.42	3.83 ± 0.45
Swin	GradMatch	10	6.06	92.54 ± 0.51	3.78 ± 0.49
Swin	GradMatch	30	11.95	94.69 ± 0.24	4.77 ± 0.38
Swin	GradMatch	50	18.23	94.91 ± 0.18	6.35 ± 0.47
Swin	GradMatch	80	40.95	95.34 ± 0.19	8.84 ± 1.00
Swin	GLISTER	1	3.08	54.85 ± 7.01	3.94 ± 0.46
Swin	GLISTER	10	3.33	92.90 ± 0.77	3.71 ± 0.42
Swin	GLISTER	30	9.22	94.39 ± 0.43	5.19 ± 0.15
Swin	GLISTER	50	15.50	94.83 ± 0.26	6.68 ± 0.79
Swin	GLISTER	80	38.22	95.34 ± 0.20	9.03 ± 1.20
Swin	Random	1	0.07	53.24 ± 7.38	3.95 ± 0.44
Swin	Random	10	0.32	92.47 ± 0.22	4.05 ± 0.46
Swin	Random	30	6.21	94.50 ± 0.36	5.10 ± 0.59
Swin	Random	50	12.50	94.68 ± 0.25	6.72 ± 0.64
Swin	Random	80	35.21	95.28 ± 0.28	8.93 ± 1.39

Table A26: Performance scores for ViT-B_16 and Swin with ImageNet-21k weights and Random initialized on Oxford-IIIT Pet.

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		239.53			-			40.35	
100	89		205.31			-			40.35	
	79		181.98			-			40.35	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	282.06	200.85	153.73	162.29	81.09	33.97	24.53	22.96	19.35
50	89	232.38	167.42	119.53	129.83	64.87	16.98	24.53	22.96	19.35
	79	204.23	139.28	107.61	113.60	48.65	16.98	24.10	22.25	19.35
	105	233.06	151.86	104.74	162.29	81.09	33.97	16.60	15.21	13.32
30	89	190.08	125.12	77.230	129.83	64.87	16.98	16.28	15.14	13.32
	79	167.32	102.36	70.69	113.60	48.65	16.98	15.96	14.60	13.32
	105	184.07	102.86	55.74	162.29	81.09	33.97	9.71	8.75	7.39
10	89	148.50	83.54	35.64	129.83	64.87	16.98	9.53	8.14	7.39
	79	130.19	65.24	33.57	113.60	48.65	16.98	8.75	7.82	7.39
	105	173.18	91.98	44.85	162.29	81.09	33.97	4.64	4.42	4.53
1	89	139.16	74.20	26.31	129.83	64.87	16.98	4.64	4.42	4.53
	79	121.69	56.74	25.07	113.60	48.65	16.98	4.64	4.42	4.53

Table A27: Performance scores for ViT-B_16 with ImageNet-21k weights on Ultramnist and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		239.53			-			40.35	
100	89		205.31			-			40.35	
	79		181.98			-			40.35	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	276.06	233.25	207.70	85.52	42.71	17.17	36.75	33.21	32.75
80	89	231.73	197.49	171.90	68.42	34.17	8.58	36.57	32.71	32.75
	79	205.04	170.80	153.75	59.87	25.63	8.58	36.10	32.60	32.75
	105	205.29	162.48	136.93	85.52	42.71	17.17	25.71	22.82	19.46
50	89	169.52	135.27	109.68	68.42	34.17	8.58	24.96	22.57	19.28
	79	150.60	116.36	99.31	59.87	25.63	8.58	24.96	21.96	19.10
	105	156.29	113.49	87.94	85.52	42.71	17.17	18.82	14.78	13.28
30	89	128.04	93.79	68.20	68.42	34.17	8.58	18.21	14.78	13.28
	79	111.71	77.47	60.43	59.87	25.63	8.58	18.21	14.46	13.07
	105	107.30	64.49	38.94	85.52	42.71	17.17	9.25	7.53	7.42
10	89	86.56	52.32	26.73	68.42	34.17	8.58	9.25	7.53	6.75
	79	75.42	41.18	24.13	59.87	25.63	8.58	7.42	7.53	6.75
	105	96.41	53.60	28.05	85.52	42.71	17.17	3.96	4.14	4.39
1	89	77.75	43.50	17.91	68.42	34.17	8.58	3.96	4.14	4.32
	79	67.95	33.71	16.67	59.87	25.63	8.58	3.96	4.14	4.28

Table A28: Performance scores for ViT-B_16 with ImageNet-21k weights on Ultramnist and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
100	105		239.53			-			40.35	
	89		205.31			-			40.35	
	79		181.98			-			40.35	
		$SSI{=}10$	$SSI{=}20$	$SSI{=}50$	$SSI{=}10$	$SSI{=}20$	$SSI{=}50$	$SSI{=}10$	$SSI{=}20$	$SSI{=}50$
80	105	337.64	268.18	220.34	147.10	77.64	29.80	37.35	35.42	32.78
	89	280.06	224.50	177.28	117.68	62.11	14.90	37.35	35.28	32.78
	79	246.69	190.30	158.62	102.97	46.58	14.90	37.35	35.28	32.78
50	105	266.86	197.41	149.57	147.10	77.64	29.80	24.35	24.28	20.14
	89	220.23	164.67	117.45	117.68	62.11	14.90	24.35	24.28	19.82
	79	193.60	137.21	105.53	102.97	46.58	14.90	24.35	24.00	19.82
30	105	217.87	148.41	100.57	147.10	77.64	29.80	19.07	15.64	13.53
	89	177.92	122.36	75.14	117.68	62.11	14.90	17.89	15.64	13.53
	79	156.68	100.30	68.61	102.97	46.58	14.90	17.39	14.57	13.53
10	105	168.87	99.42	51.58	147.10	77.64	29.80	5.00	7.46	6.92
	89	136.34	80.78	33.56	117.68	62.11	14.90	5.00	7.46	6.92
	79	119.56	63.17	31.49	102.97	46.58	14.90	5.00	7.46	6.75
1	105	157.99	88.53	40.69	147.10	77.64	29.80	4.14	3.96	4.21
	89	127.01	71.45	24.23	117.68	62.11	14.90	4.14	3.96	4.21
	79	111.06	54.67	22.99	102.97	46.58	14.90	4.14	3.96	4.21

Table A29: Performance scores for ViT-B_16 with ImageNet-21k weights on Ultramnist and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
100	105		239.53			-			40.35	
	89		205.31			-			40.35	
	79		181.98			-			40.35	
		$SSI{=}10$	$SSI{=}20$	$SSI{=}50$	$SSI{=}10$	$SSI{=}20$	$SSI{=}50$	$SSI{=}10$	$SSI{=}20$	$SSI{=}50$
80	105	190.53	190.53	190.53	0.00	0.00	0.00	-	33.89	-
	89	162.38	162.38	162.38	0.00	0.00	0.00	-	33.89	-
	79	143.71	143.71	143.71	0.00	0.00	0.00	-	33.89	-
50	105	119.76	119.76	119.76	0.00	0.00	0.00	18.39	18.92	19.28
	89	102.55	102.55	102.55	0.00	0.00	0.00	18.39	18.92	19.28
	79	90.62	90.62	90.62	0.00	0.00	0.00	18.39	18.92	19.25
30	105	70.77	70.77	70.77	0.00	0.00	0.00	12.17	13.03	12.75
	89	60.24	60.24	60.24	0.00	0.00	0.00	12.17	13.03	12.75
	79	53.71	53.71	53.71	0.00	0.00	0.00	12.17	13.03	12.75
10	105	21.77	21.77	21.77	0.00	0.00	0.00	7.03	7.50	7.75
	89	18.66	18.66	18.66	0.00	0.00	0.00	7.03	7.50	7.75
	79	16.59	16.59	16.59	0.00	0.00	0.00	7.03	7.50	7.75
1	105	10.88	10.88	10.88	0.00	0.00	0.00	4.00	3.96	4.21
	89	9.33	9.33	9.33	0.00	0.00	0.00	4.00	3.96	4.21
	79	8.08	8.08	8.08	0.00	0.00	0.00	4.00	3.96	4.21

Table A30: Performance scores for ViT-B_16 with ImageNet-21k weights on UltraMNIST and its Effective data per epoch (EDPE), method=Random

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		114.99			-			61.28	
100	89		98.56			-			61.28	
	79		87.36			-			61.21	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	138.70	115.05	100.78	47.23	23.58	9.31	54.28	51.78	48.21
80	89	115.74	96.82	82.61	37.78	18.86	4.65	53.60	51.75	47.39
	79	102.05	83.14	73.65	33.06	14.15	4.65	52.07	50.85	46.92
	105	104.72	81.07	66.80	47.23	23.58	9.31	40.28	34.25	28.82
50	89	87.01	68.09	53.88	37.78	18.86	4.65	40.28	34.25	28.67
	79	76.56	57.65	48.16	33.06	14.15	4.65	39.89	32.71	28.32
	105	81.20	57.55	43.28	47.23	23.58	9.31	24.57	21.39	17.39
30	89	66.70	47.78	33.57	37.78	18.86	4.65	24.46	21.39	17.28
	79	58.84	39.93	30.44	33.06	14.15	4.65	24.46	20.67	17.28
	105	57.68	34.03	19.76	47.23	23.58	9.31	15.78	13.00	11.46
10	89	46.74	27.82	13.61	37.78	18.86	4.65	15.78	13.00	11.46
	79	41.02	22.11	12.62	33.06	14.15	4.65	15.78	13.00	11.46
	105	52.45	28.81	14.54	47.23	23.58	9.31	6.92	5.50	5.85
1	89	42.26	23.34	9.13	37.78	18.86	4.65	6.78	5.50	5.60
	79	36.94	18.03	8.53	33.06	14.15	4.65	6.25	5.46	5.39

Table A31: Performance scores for ResNet50 with ImageNet-21k weights on Ul-traMNIST and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		114.99			-			61.28	
100	89		98.56			-			61.28	
	79		87.36			-			61.21	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	126.53	109.21	98.26	35.06	17.74	6.79	52.71	51.46	49.03
80	89	105.20	91.35	80.55	28.05	14.19	3.39	52.71	51.39	48.96
	79	94.23	80.33	73.08	24.54	10.64	3.39	52.57	50.82	48.82
	105	92.55	75.23	64.29	35.06	17.74	6.79	42.17	38.07	32.32
50	89	75.34	61.48	50.68	28.05	14.19	3.39	41.64	37.64	31.28
	79	66.85	52.95	45.71	24.54	10.64	3.39	41.46	36.67	31.28
	105	69.03	51.71	40.76	35.06	17.74	6.79	28.92	22.46	18.89
30	89	55.42	41.57	30.77	28.05	14.19	3.39	28.10	21.60	18.89
	79	49.43	35.53	28.28	24.54	10.64	3.39	26.17	21.10	18.89
	105	45.51	28.19	17.24	35.06	17.74	6.79	15.10	13.28	10.10
10	89	38.00	24.14	13.35	28.05	14.19	3.39	15.10	13.28	10.10
	79	32.01	18.11	10.86	24.54	10.64	3.39	14.82	13.03	10.10
	105	40.29	22.96	12.02	35.06	17.74	6.79	6.53	6.67	5.10
1	89	32.53	18.67	7.87	28.05	14.19	3.39	6.53	6.67	5.10
	79	28.42	14.52	7.28	24.54	10.64	3.39	6.53	5.96	5.00

Table A32: Performance scores for ResNet50 with ImageNet-21k weights on Ul-traMNIST and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		114.99			-			61.28	
100	89		98.56			-			61.28	
	79		87.36			-			61.21	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	142.94	111.20	100.78	51.47	19.73	6.78	53.92	53.00	-
80	89	119.13	93.74	82.61	41.18	15.78	3.39	53.92	52.67	-
	79	105.02	80.83	73.65	36.03	11.84	3.39	53.92	50.82	-
	105	108.97	77.23	64.28	51.47	19.73	6.78	40.53	34.85	31.00
50	89	90.41	65.02	52.62	41.18	15.78	3.39	40.53	34.82	31.00
	79	79.54	55.35	46.90	36.03	11.84	3.39	39.53	34.35	31.00
	105	85.45	53.71	40.76	51.47	19.73	6.78	28.25	22.78	20.71
30	89	70.10	44.71	32.31	41.18	15.78	3.39	28.25	22.78	20.42
	79	61.81	37.62	29.18	36.03	11.84	3.39	27.25	11.84	20.42
	105	61.92	30.19	17.24	51.47	19.73	6.78	17.50	13.64	11.46
10	89	50.14	24.75	12.35	41.18	15.78	3.39	16.85	13.64	11.46
	79	43.99	19.80	11.35	36.03	11.84	3.39	16.85	13.64	11.46
	105	56.70	24.96	12.01	51.47	19.73	6.78	9.50	7.82	6.03
1	89	45.66	20.27	7.87	41.18	15.78	3.39	9.50	7.67	6.03
	79	39.91	15.72	7.27	36.03	11.84	3.39	8.42	7.25	6.03

Table A33: Performance scores for ResNet50 with ImageNet-21k weights on UltraMNIST and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Accuracy		
	105		239.53			-			40.35	
100	89		205.31			-			40.35	
	79		181.98			-			40.35	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	91.47	91.47	91.47	0.00	0.00	0.00	48.75	49.35	49.75
80	89	77.95	77.95	77.95	0.00	0.00	0.00	48.71	49.35	49.71
	79	68.99	68.99	68.99	0.00	0.00	0.00	48.42	49.32	49.60
	105	57.49	57.49	57.49	0.00	0.00	0.00	30.25	28.75	29.75
50	89	49.23	49.23	49.23	0.00	0.00	0.00	30.25	28.75	29.71
	79	43.50	43.50	43.50	0.00	0.00	0.00	30.25	28.75	29.67
	105	33.97	33.97	33.97	0.00	0.00	0.00	20.07	19.17	19.14
30	89	28.92	28.92	28.92	0.00	0.00	0.00	20.07	19.17	19.14
	79	25.78	25.78	25.78	0.00	0.00	0.00	20.07	19.17	19.14
	105	10.45	10.45	10.45	0.00	0.00	0.00	12.46	12.89	11.89
10	89	8.96	8.96	8.96	0.00	0.00	0.00	12.46	12.89	11.89
	79	7.96	7.96	7.96	0.00	0.00	0.00	12.46	12.89	11.89
	105	5.22	5.22	5.22	0.00	0.00	0.00	5.75	6.89	5.67
1	89	4.48	4.48	4.48	0.00	0.00	0.00	5.75	6.89	5.67
	79	3.88	3.88	3.88	0.00	0.00	0.00	5.64	6.82	5.53

Table A34: Performance scores for ResNet50 with ImageNet-21k weights on UltraMNIST and its Effective data per epoch (EDPE), method=Random

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
100	105		16.59			-			0.88	
	89		14.22			-			0.88	
	79		12.64			-			0.88	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
80	105	15.91	14.88	14.25	2.08	1.05	0.42	0.87	0.89	0.90
	89	13.52	12.69	12.06	1.67	0.84	0.21	0.87	0.89	0.90
	79	11.99	11.16	10.74	1.46	0.63	0.21	0.87	0.89	0.90
50	105	10.38	9.35	8.72	2.08	1.05	0.42	0.87	0.86	0.88
	89	8.78	7.95	7.32	1.67	0.84	0.21	0.87	0.86	0.88
	79	7.78	6.95	6.53	1.46	0.63	0.21	0.87	0.86	0.88
30	105	7.61	6.58	5.95	2.08	1.05	0.42	0.86	0.86	0.87
	89	6.41	5.58	4.95	1.67	0.84	0.21	0.86	0.86	0.87
	79	5.67	4.84	4.42	1.46	0.63	0.21	0.86	0.86	0.87
10	105	4.16	3.12	2.50	2.08	1.05	0.42	0.86	0.84	0.85
	89	3.44	2.62	1.99	1.67	0.84	0.21	0.85	0.83	0.85
	79	3.04	2.21	1.79	1.46	0.63	0.21	0.85	0.83	0.85
1	105	2.26	1.22	0.60	2.08	1.05	0.42	0.73	0.72	0.66
	89	1.81	0.99	0.36	1.67	0.84	0.21	0.73	0.72	0.66
	79	1.59	0.76	0.34	1.46	0.63	0.21	0.71	0.72	0.66

Table A35: Performance scores for ViT_B16 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
100	105		16.59			-			0.88	
	89		14.22			-			0.88	
	79		12.64			-			0.88	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
80	105	15.61	14.83	14.20	1.79	1.00	0.37	0.88	0.88	0.88
	89	13.28	12.65	12.04	1.43	0.80	0.18	0.88	0.88	0.88
	79	11.78	11.13	10.72	1.25	0.60	0.18	0.88	0.88	0.88
50	105	10.08	9.30	8.67	1.79	1.00	0.37	0.89	0.88	0.88
	89	8.54	7.91	7.30	1.43	0.80	0.18	0.89	0.88	0.88
	79	7.57	6.92	6.51	1.25	0.60	0.18	0.89	0.88	0.88
30	105	7.32	6.53	5.90	1.79	1.00	0.37	0.88	0.87	0.88
	89	6.17	5.54	4.93	1.43	0.80	0.18	0.88	0.87	0.88
	79	5.46	4.81	4.40	1.25	0.60	0.18	0.88	0.87	0.88
10	105	3.86	3.08	2.45	1.79	1.00	0.37	0.85	0.84	0.85
	89	3.21	2.58	1.96	1.43	0.80	0.18	0.85	0.84	0.85
	79	2.83	2.18	1.76	1.25	0.60	0.18	0.85	0.84	0.85
1	105	1.96	1.17	0.55	1.79	1.00	0.37	0.66	0.74	0.70
	89	1.58	0.95	0.33	1.43	0.80	0.18	0.66	0.74	0.70
	79	1.38	0.73	0.32	1.25	0.60	0.18	0.66	0.74	0.70

Table A36: Performance scores for ViT_B16 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
			16.59			-			0.88	
100	89		14.22			-			0.88	
	79		12.64			-			0.88	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	16.46	15.77	14.51	3.41	1.94	0.68	0.88	0.88	0.88
80	89	13.95	13.40	12.19	2.73	1.55	0.34	0.88	0.88	0.88
	79	12.37	11.70	10.87	2.38	1.16	0.34	0.88	0.88	0.88
	105	11.71	10.24	8.98	3.41	1.94	0.68	0.87	0.88	0.87
50	89	9.84	8.66	7.45	2.73	1.55	0.34	0.87	0.88	0.87
	79	8.71	7.48	6.66	2.38	1.16	0.34	0.87	0.88	0.87
	105	8.94	7.47	6.22	3.41	1.94	0.68	0.86	0.87	0.87
30	89	7.47	6.29	5.08	2.73	1.55	0.34	0.86	0.87	0.87
	79	6.60	5.38	4.55	2.38	1.16	0.34	0.86	0.87	0.87
	105	5.48	4.01	2.76	3.41	1.94	0.68	0.75	0.76	0.81
10	89	4.50	3.33	2.12	2.73	1.55	0.34	0.75	0.76	0.81
	79	3.96	2.74	1.92	2.38	1.16	0.34	0.75	0.76	0.81
	105	3.58	2.11	0.86	3.41	1.94	0.68	0.70	0.67	0.54
1	89	2.87	1.70	0.49	2.73	1.55	0.34	0.58	0.67	0.54
	79	2.52	1.29	0.47	2.38	1.16	0.34	0.58	0.67	0.54

Table A37: Performance scores for ViT_B16 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
	105		16.59			-			0.88	
100	89		14.22			-			0.88	
	79		12.64			-			0.88	
		SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50	SSI=10	SSI=20	SSI=50
	105	13.82	13.82	13.82	0.0	0.0	0.0	0.88	0.88	0.89
80	89	11.85	11.85	11.85	0.0	0.0	0.0	0.88	0.88	0.89
	79	10.53	10.53	10.53	0.0	0.0	0.0	0.88	0.88	0.89
	105	8.29	8.29	8.29	0.0	0.0	0.0	0.88	0.88	0.87
50	89	7.11	7.11	7.11	0.0	0.0	0.0	0.88	0.88	0.87
	79	6.32	6.32	6.32	0.0	0.0	0.0	0.88	0.88	0.87
	105	5.53	5.53	5.53	0.0	0.0	0.0	0.87	0.87	0.87
30	89	4.74	4.74	4.74	0.0	0.0	0.0	0.87	0.87	0.87
	79	4.21	4.21	4.21	0.0	0.0	0.0	0.87	0.87	0.87
	105	2.07	2.07	2.07	0.0	0.0	0.0	0.83	0.83	0.86
10	89	1.77	1.77	1.77	0.0	0.0	0.0	0.83	0.83	0.86
	79	1.58	1.58	1.58	0.0	0.0	0.0	0.83	0.83	0.85
	105	0.17	0.17	0.17	0.0	0.0	0.0	0.69	0.66	0.75
1	89	0.14	0.14	0.14	0.0	0.0	0.0	0.69	0.66	0.75
	79	0.13	0.13	0.13	0.0	0.0	0.0	0.69	0.66	0.75

Table A38: Performance scores for ViT_B16 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=Random

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
	105		10.73			-			0.88	
100	89		9.20			-			0.88	
	79		8.18			-			0.88	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	10.76	9.86	9.32	1.81	0.91	0.37	0.90	0.90	0.90
80	89	9.12	8.40	7.85	1.45	0.73	0.18	0.90	0.89	0.90
	79	8.08	7.37	7.00	1.26	0.55	0.18	0.89	0.89	0.90
	105	7.18	6.28	5.74	1.81	0.91	0.37	0.88	0.88	0.88
50	89	6.05	5.33	4.79	1.45	0.73	0.18	0.88	0.88	0.88
	79	5.36	4.64	4.27	1.26	0.55	0.18	0.87	0.88	0.88
	105	5.39	4.49	3.95	1.81	0.91	0.37	0.87	0.87	0.88
30	89	4.51	3.80	3.25	1.45	0.73	0.18	0.83	0.85	0.88
	79	3.99	3.27	2.91	1.26	0.55	0.18	0.83	0.84	0.87
	105	3.15	2.26	1.71	1.81	0.91	0.37	0.86	0.82	0.84
10	89	2.60	1.88	1.33	1.45	0.73	0.18	0.85	0.82	0.84
	79	2.29	1.57	1.21	1.26	0.55	0.18	0.85	0.82	0.84
	105	1.92	1.03	0.48	1.81	0.91	0.37	0.81	0.74	0.84
1	89	1.54	0.83	0.28	1.45	0.73	0.18	0.81	0.71	0.83
	79	1.35	0.63	0.27	1.26	0.55	0.18	0.79	0.69	0.83

Table A39: Performance scores for ResNet50 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=CRAIG

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
	105		10.73			-			0.88	
100	89		9.20			-			0.88	
	79		8.18			-			0.88	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	10.46	9.71	9.26	1.51	0.76	0.31	0.90	0.89	0.89
80	89	8.88	8.28	7.82	1.21	0.61	0.15	0.90	0.89	0.89
	79	7.88	7.27	6.97	1.06	0.45	0.15	0.90	0.89	0.89
	105	6.88	6.13	5.68	1.51	0.76	0.31	0.88	0.89	0.89
50	89	5.81	5.21	4.75	1.21	0.61	0.15	0.88	0.89	0.89
	79	5.15	4.54	4.24	1.06	0.45	0.15	0.88	0.89	0.89
	105	5.09	4.34	3.89	1.51	0.76	0.31	0.88	0.88	0.88
30	89	4.28	3.67	3.22	1.21	0.61	0.15	0.88	0.88	0.88
	79	3.79	3.18	2.88	1.06	0.45	0.15	0.87	0.88	0.88
	105	2.86	2.10	1.65	1.51	0.76	0.31	0.86	0.87	0.85
10	89	2.36	1.76	1.30	1.21	0.61	0.15	0.86	0.87	0.85
	79	2.08	1.48	1.17	1.06	0.45	0.15	0.86	0.86	0.85
	105	1.62	0.87	0.42	1.51	0.76	0.31	0.79	0.80	0.73
1	89	1.31	0.70	0.25	1.21	0.61	0.15	0.79	0.80	0.73
	79	1.14	0.54	0.24	1.06	0.45	0.15	0.79	0.80	0.73

Table A40: Performance scores for ResNet50 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=GradMatch

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
	105		10.73			-			0.88	
100	89		9.20			-			0.88	
	79		8.18			-			0.88	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	12.46	10.70	9.65	3.51	1.76	0.70	0.90	0.90	0.89
80	89	10.48	9.07	8.02	2.81	1.40	0.35	0.90	0.90	0.89
	79	9.28	7.87	7.17	2.46	1.05	0.35	0.89	0.89	0.89
	105	8.88	7.13	6.07	3.51	1.76	0.70	0.88	0.88	0.87
50	89	7.41	6.01	4.95	2.81	1.40	0.35	0.86	0.88	0.87
	79	6.55	5.14	4.44	2.46	1.05	0.35	0.86	0.85	0.87
	105	7.09	5.34	4.28	3.51	1.76	0.70	0.73	0.85	0.88
30	89	5.88	4.47	3.42	2.81	1.40	0.35	0.73	0.85	0.88
	79	5.18	3.78	3.08	2.46	1.05	0.35	0.73	0.85	0.88
	105	4.85	3.10	2.05	3.51	1.76	0.70	0.78	0.81	0.85
10	89	3.96	2.55	1.50	2.81	1.40	0.35	0.78	0.81	0.83
	79	3.48	2.07	1.37	2.46	1.05	0.35	0.78	0.80	0.83
	105	3.62	1.87	0.82	3.51	1.76	0.70	0.64	0.68	0.74
1	89	2.90	1.50	0.45	2.81	1.40	0.35	0.64	0.62	0.74
	79	2.54	1.14	0.43	2.46	1.05	0.35	0.64	0.62	0.74

Table A41: Performance scores for ResNet50 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=GLISTER

EDPE (%)	Epoch	Total Time (Mins)			Total Selection Time (Mins)			Quadratic κ		
	105		10.73			-			0.88	
100	89		9.20			-			0.88	
	79		8.18			-			0.88	
		SSI=10	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*	*SSI=10*	*SSI=20*	*SSI=50*
	105	8.94	8.94	8.94	0.0	0.0	0.0	0.89	0.89	0.89
80	89	7.67	7.67	7.67	0.0	0.0	0.0	0.89	0.89	0.89
	79	6.81	6.81	6.81	0.0	0.0	0.0	0.89	0.89	0.89
	105	5.36	5.36	5.36	0.0	0.0	0.0	0.88	0.89	0.88
50	89	4.60	4.60	4.60	0.0	0.0	0.0	0.88	0.89	0.88
	79	4.09	4.09	4.09	0.0	0.0	0.0	0.88	0.89	0.88
	105	3.57	3.57	3.57	0.0	0.0	0.0	0.87	0.87	0.87
30	89	3.06	3.06	3.06	0.0	0.0	0.0	0.87	0.87	0.87
	79	2.72	2.72	2.72	0.0	0.0	0.0	0.87	0.87	0.87
	105	1.34	1.34	1.34	0.0	0.0	0.0	0.84	0.86	0.86
10	89	1.15	1.15	1.15	0.0	0.0	0.0	0.84	0.85	0.86
	79	1.02	1.02	1.02	0.0	0.0	0.0	0.84	0.85	0.86
	105	0.11	0.11	0.11	0.0	0.0	0.0	0.82	0.79	0.78
1	89	0.09	0.09	0.09	0.0	0.0	0.0	0.82	0.79	0.78
	79	0.08	0.08	0.08	0.0	0.0	0.0	0.82	0.79	0.78

Table A42: Performance scores for ResNet50 with ImageNet-21k weights on Medical Dataset APTOS-2019 and its Effective data per epoch (EDPE), method=Random