BiasPruner: Debiased Continual Learning for Medical Image Classification (Supplementary Material)

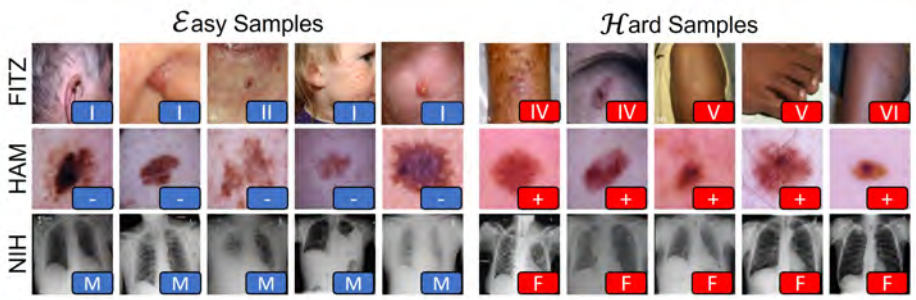

Fig. 4: Visualization of easy (blue square) and hard (red square) samples across the different benchmarks. 1st row: Images from FITZ (Task 4), where each image is labeled by its Fitzpatrick skin tone, denoted as I, II, IV, V or VI. 2nd row: Images from HAM (Task 3), where each image is labeled by age (age $<$ or ≥ 60), denoted as $-$ or $+$, respectively. 3rd row: Images from NIH (Task 1), where each image is labeled as male or female, denoted as M or F, respectively. We notice that the hard samples represent the minority within their respective tasks.

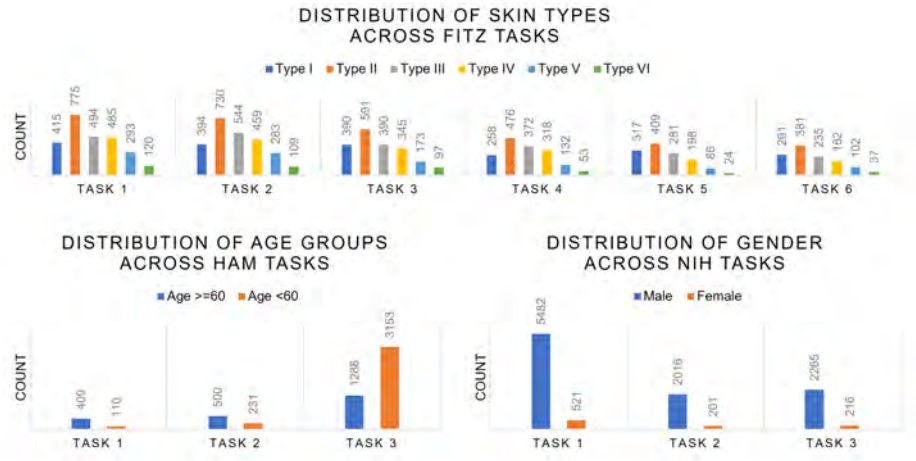

Fig. 5: Bias distribution across the different tasks in FITZ, HAM and NIH.

Table 5: Distribution of images across train, validation, and test sets for each task in FITZ, HAM, and NIH. The V value between brackets next to Train represents the Cramer's V correlation between the sensitive attribute (e.g., skin tone in FITZ) and disease classes.

	FITZ						**HAM**			**NIH**		
	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 1	Task 2	Task 3	Task 1	Task 2	Task 3
Train (V)	2,582 (0.311)	2,519 (0.287)	1,986 (0.277)	1,609 (0.333)	1,315 (0.293)	1,198 (0.328)	519 (0.187)	731 (0.112)	4,441 (0.199)	6,003 (0.259)	2,217 (0.242)	2,481 (0.264)
Val	361	358	286	230	180	171	85	123	809	1,711	646	710
Test	747	721	568	462	375	344	167	240	1,563	3,472	1,311	1,442
Total	3,690	3,598	2,840	2,301	1,870	1,713	771	1,094	6,813	11,186	4,174	4,633

Table 6: Continuation of Table 2 –Standard deviation results.

Exp	Method	F	ACC							DPR	EOD
			Type-I	Type-II	Type-III	Type-IV	Type-V	Type-VI	Overall		
Comparison against Baselines											
$\mathcal{A}$	JOINT	0.81	0.28	0.06	0.85	0.28	0.35	0.89	0.69	0.66	0.51
	SINGLE	0.73	0.13	0.15	0.30	0.45	0.53	0.19	0.11	0.12	0.38
	SeqFT	2.24	1.68	2.11	2.48	2.34	2.58	2.64	2.33	2.56	2.84
Comparison against CL Methods											
$\mathcal{B}$	EWC	1.55	1.03	1.19	1.15	1.83	1.78	1.21	1.11	1.22	1.05
	PackNet	0.21	0.57	0.41	0.93	0.25	0.81	0.35	0.69	0.76	0.49
	SupSup	0.98	0.66	0.52	0.63	0.55	0.82	0.54	0.44	0.38	0.79
Comparison against CL with Bias Mitigation Methods											
$\mathcal{C}$	EWC+S	1.51	1.76	1.34	1.19	1.75	1.26	1.33	1.28	1.01	1.14
	PackNet+S	0.15	0.14	0.62	0.51	0.26	0.17	0.64	0.25	0.46	0.26
	SupSup+S	0.25	0.58	0.61	0.63	0.24	0.5	0.28	0.56	0.63	0.65
	EWC+W	1.37	1.27	1.44	1.68	1.34	1.25	1.42	1.62	2.53	1.48
	PackNet+W	0.28	0.34	0.58	0.35	0.22	0.23	0.31	0.37	0.34	0.71
	SupSup+W	0.63	0.94	0.41	0.37	0.82	0.21	0.52	0.16	0.35	0.81
Our Proposed Fair Continual Learning Method											
$\mathcal{D}$	`BiasPruner`	0.54	0.38	0.15	0.84	0.44	0.39	0.71	0.24	0.33	0.41
[Upper-bound] Comparison against a Bias Mitigation Method											
$\mathcal{E}$	FairDisCo	0.91	0.22	0.84	0.65	0.74	0.33	0.85	0.12	0.35	0.34

Table 7: Continuation of Table 3 –Standard deviation results.

Exp	Method	HAM						NIH					
		F	ACC			DPR	EOD	F	ACC			DPR	EOD
			<60	≥60	Overall				M	F	Overall		
Comparison against Baselines													
$\mathcal{F}$	JOINT	0.05	0.14	0.11	0.14	0.09	0.08	0.15	0.08	0.11	0.22	0.61	0.23
	SINGLE	0.53	0.72	0.23	0.91	0.34	0.82	0.91	0.79	0.44	0.27	0.22	0.42
	SeqFT	2.64	0.65	0.85	0.36	0.41	0.45	0.54	0.49	0.96	0.46	0.22	0.36
Comparison against CL Methods													
$\mathcal{G}$	EWC	1.46	1.12	1.69	1.79	1.14	1.16	1.74	1.62	1.42	1.66	1.19	1.77
	PackNet	0.44	0.57	0.28	0.22	0.82	0.61	0.22	0.43	0.51	0.17	0.34	0.55
	SupSup	0.53	0.64	0.22	0.74	0.91	0.84	0.39	0.53	0.55	0.69	0.22	0.43
Comparison against CL with Bias Mitigation Methods													
$\mathcal{H}$	EWC+S	1.64	1.83	1.72	1.11	1.96	1.66	1.36	1.19	1.57	1.21	1.44	1.36
	PackNet+S	0.43	0.37	0.64	0.87	0.74	0.22	0.42	0.33	0.35	0.19	0.43	0.35
	SupSup+S	0.18	0.44	0.61	0.86	0.34	0.63	0.91	0.53	0.28	0.39	0.38	0.19
	EWC+W	1.17	1.75	1.81	1.62	1.56	1.38	1.35	1.86	1.54	1.81	1.71	1.26
	PackNet+W	0.84	0.35	0.71	0.23	0.55	0.14	0.45	0.82	0.22	0.65	0.27	0.33
	SupSup+W	0.72	0.43	0.16	0.44	0.78	0.81	0.62	0.77	0.25	0.34	0.64	0.81
Our Proposed Fair Continual Learning Method													
$\mathcal{I}$	`BiasPruner`	0.44	0.18	0.57	0.21	0.38	0.25	0.35	0.41	0.27	0.39	0.44	0.62
[Upper-bound] Comparison against a Bias Mitigation Method													
$\mathcal{J}$	FairDisCo	0.92	0.64	0.38	0.39	0.22	0.37	0.72	0.21	0.34	0.61	0.66	0.74