BiomedHELIX : HiErarchical-Local Interaction eXploration for biomedical vision-language models

Supplementary Material

6. Dataset Details

Table 6 provides detailed statistics for the 11 datasets used in our experiments. It spans 9 imaging modalities and 10 organs, summarizing the specific class names and the exact number of samples in the train, validation, and test splits.

7. Additional Few-shot Results

Table 7 provides detailed comparisons of BiomedHELIX and prior state-of-the-art methods across 11 biomedical datasets for $K \in \{4, 8, 16\}$. Consistent with the analysis in the main text, BiomedHELIX displays a performance advantage that becomes increasingly pronounced as data availability grows. While performing comparably to the strong BiomedCoOp baseline in the data-starved regime ($K = 4$), BiomedHELIX establishes a decisive lead at $K = 16$. This dataset-level breakdown confirms our central thesis: by explicitly unlocking hierarchical and local visual information, BiomedHELIX overcomes the "visual bottleneck" inherent in methods limited to global features, granting it a structurally superior capacity to learn from increased visual evidence.

8. Effect of Context Length

Table 4 investigates the impact of context length (M) on model performance. We observe that the performance initially improves as the context length increases from 2 to 4, achieving the optimal trade-off at $M = 4$ with the highest Harmonic Mean (HM) of 76.81%. Notably, $M = 4$ yields a significant gain of 7.1% in novel class accuracy compared to $M = 2$. However, further extending the context length to 8, 16, or 32 leads to a drastic performance collapse, with HM scores dropping significantly to around 51%. This sharp decline indicates that while a moderate context length is beneficial for representation learning, an excessively long context may introduce optimization difficulties or overfitting, thereby severely diminishing the model's generalization ability.

9. Additional Hyperparameters

Table 5 presents the dataset-specific hyperparameters (α, β, and λ) optimized for BiomedHELIX across both few-shot learning and base-to-novel generalization benchmarks. Specifically, α and β modulate the influence of the losses l_H and l_G, respectively, whereas λ dictates the magnitude of the regularization term. We observe that while α and β exhibit significant variance to accommodate diverse dataset distributions, λ consistently remains at a low range (typically ≤ 0.05) to ensure training stability.

Table 4. Effect of the context vector length on classification accuracy (%) in Base-to-Novel generalization.

Context Length	Base Acc.	Novel Acc.	HM
2	79.24	66.39	72.25
4	80.44	73.49	76.81
8	58.76	45.31	51.17
16	56.37	46.06	50.70
32	56.72	47.18	51.51

Table 5. Hyperparameter values for α, β, and λ across different datasets and benchmarks.

Dataset	Benchmark	α	β	λ
BTMRI	Few-shot	0.3	0.7	0.01
	Base-to-Novel	0.3	0.2	0.01
BUSI	Few-shot	0.9	0.9	0.2
	Base-to-Novel	-	-	-
COVID-QU-Ex	Few-shot	0.9	0.3	0.1
	Base-to-Novel	0.2	0.4	0.1
CTKIDNEY	Few-shot	0.1	0.1	0.01
	Base-to-Novel	0.3	0.2	0.05
DermaMNIST	Few-shot	0.3	0.7	0.01
	Base-to-Novel	0.9	0.8	0.1
Kvasir	Few-shot	0.9	0.9	0.05
	Base-to-Novel	0.5	0.2	0.01
CHMNIST	Few-shot	0.1	0.7	0.01
	Base-to-Novel	0.4	0.7	0.15
LC25000	Few-shot	0.9	0.1	0.05
	Base-to-Novel	0.8	0.1	0.3
RETINA	Few-shot	0.7	0.9	0.01
	Base-to-Novel	0.5	0.2	0.01
KneeXray	Few-shot	0.7	0.3	0.0
	Base-to-Novel	0.1	0.9	0.01
OCTMNIST	Few-shot	0.7	0.7	0.02
	Base-to-Novel	0.7	0.1	0.01

Table 6. An overview of the 11 datasets used spanning 9 biomedical imaging modalities and 10 different organs.

Dataset	Modality	Organ	Train	Val	Test	Classes
CTKidney	CT	Kidney	6,221	2,487	3,738	Kidney Cyst, Kidney Stone, Kidney Tumor, Normal Kidney
DermaMNIST	Dermatoscopy	Skin	7,007	1,003	2,005	Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis, Dermatofibroma, Melanocytic nevus, Melanoma, Vascular Lesion
Kvasir	Endoscopy	Colon	2,000	800	1,200	Dyed Lifted Polyps, Normal Cecum, Esophagitis, Dyed Resection Margins, Normal Pylorus, Normal Z Line, Polyps, Ulcerative Colitis
RETINA	Fundus Photo.	Retina	2,108	841	1,268	Cataract, Diabetic Retinopathy, Glaucoma, Normal Retina
LC25000	Histopathology	Lung/Colon	12,500	5,000	7,500	Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, Lung Squamous Cell Carcinoma
CHMNIST	Histopathology	Colorectal	2,496	1,000	1,504	Adipose Tissue, Complex Stroma, Debris, Empty Background, Immune Cells, Normal Mucosal Glands, Simple Stroma, Tumor Epithelium
BTMRI	MRI	Brain	2,854	1,141	1,717	Glioma Tumor, Meningioma Tumor, Normal Brain, Pituitary Tumor
OCTMNIST	OCT	Retina	97,477	10,832	1,000	Choroidal Neovascularization, Drusen, Diabetic Macular Edema, Normal
BUSI	Ultrasound	Breast	389	155	236	Benign Tumors, Malignant Tumors, Normal Scans
COVID-QU-Ex	X-Ray	Chest	10,582	4,232	6,351	COVID-19, Lung Opacity, Normal Lungs, Viral Pneumonia
KneeXray	X-Ray	Knee	5,778	826	1,656	Osteoarthritis (No, Doubtful, Minimal, Moderate, Severe)

Table 7. Per-dataset performance comparison of BiomedHELIX (Ours) with various methods in few-shot setting ($K = 4, 8, 16$). Accuracy is reported in %.

Dataset	**Method**	$K = 4$	$K = 8$	$K = 16$
BTMRI	BiomedCLIP		56.79	
	BiomedCLIP + Ensemble		61.04	
	CLIP-Adapter	56.80	57.15	60.16
	Tip-Adapter	76.37	73.75	78.97
	Tip-Adapter-F	77.90	79.18	82.27
	Standard LP	75.98	77.63	81.24
	LP++	75.48	77.11	81.61
	CoOp	74.68	79.27	82.37
	CoCoOp	67.83	71.69	78.45
	KgCoOp	75.40	79.79	81.07
	ProGrad	76.24	78.82	82.84
	BiomedCoOp	77.23	78.55	83.30
	BiomedHELIX (Ours)	76.83	80.80	85.90
BUSI	BiomedCLIP		59.75	
	BiomedCLIP + Ensemble		59.75	
	CLIP-Adapter	61.72	61.86	63.55
	Tip-Adapter	59.03	55.93	68.78
	Tip-Adapter-F	64.54	68.50	71.89
	Standard LP	53.38	65.53	68.78
	LP++	60.31	66.10	70.05
	CoOp	60.17	64.69	69.49
	CoCoOp	59.75	65.82	70.20
	KgCoOp	62.01	67.37	70.62
	ProGrad	62.29	64.83	71.47
	BiomedCoOp	59.32	63.27	70.34
	BiomedHELIX (Ours)	62.57	67.37	75.70
COVID-QU-Ex	BiomedCLIP		43.80	
	BiomedCLIP + Ensemble		66.86	
	CLIP-Adapter	46.28	48.68	49.55
	Tip-Adapter	63.84	66.77	73.05
	Tip-Adapter-F	69.97	69.89	76.07
	Standard LP	60.55	68.29	71.98
	LP++	62.32	66.19	72.79
	CoOp	67.03	74.66	76.37
	CoCoOp	63.70	69.36	74.52
	KgCoOp	65.91	74.86	75.65
	ProGrad	68.56	74.65	74.93
	BiomedCoOp	73.28	76.26	78.72
	BiomedHELIX (Ours)	67.63	74.77	78.33

Table 7 (continued): Per-dataset performance comparison.

Dataset	Method	$K = 4$	$K = 8$	$K = 16$
CTKIDNEY	BiomedCLIP		42.43	
	BiomedCLIP + Ensemble		56.82	
	CLIP-Adapter	42.19	44.64	47.28
	Tip-Adapter	55.33	69.89	73.38
	Tip-Adapter-F	60.18	75.24	82.07
	Standard LP	69.54	78.89	82.50
	LP++	65.73	77.06	79.07
	CoOp	68.12	77.40	83.52
	CoCoOp	61.07	73.93	77.70
	KgCoOp	68.68	77.43	77.67
	ProGrad	67.90	78.23	81.13
	BiomedCoOp	66.50	77.16	83.20
	BiomedHELIX (Ours)	72.00	83.67	89.73
DermaMNIST	BiomedCLIP		38.75	
	BiomedCLIP + Ensemble		53.62	
	CLIP-Adapter	34.97	34.28	29.02
	Tip-Adapter	47.31	61.67	62.67
	Tip-Adapter-F	50.44	43.87	53.86
	Standard LP	49.77	51.02	55.34
	LP++	36.29	45.78	50.98
	CoOp	43.71	46.80	51.07
	CoCoOp	25.29	40.42	40.97
	KgCoOp	35.35	38.79	36.59
	ProGrad	43.69	51.07	46.33
	BiomedCoOp	60.07	61.98	62.59
	BiomedHELIX (Ours)	50.53	54.87	62.67
Kvasir	BiomedCLIP		54.58	
	BiomedCLIP + Ensemble		57.50	
	CLIP-Adapter	54.83	56.08	56.50
	Tip-Adapter	69.61	69.13	74.22
	Tip-Adapter-F	69.94	75.86	78.00
	Standard LP	72.38	78.88	79.00
	LP++	69.36	72.52	75.41
	CoOp	70.78	77.14	77.88
	CoCoOp	68.94	72.92	75.22
	KgCoOp	68.28	72.05	72.95
	ProGrad	70.00	76.03	75.88
	BiomedCoOp	74.08	77.72	78.89
	BiomedHELIX (Ours)	75.33	80.33	84.23

Table 7 (continued): Per-dataset performance comparison.

Dataset	Method	$K = 4$	$K = 8$	$K = 16$
CHMNIST	BiomedCLIP		30.65	
	BiomedCLIP + Ensemble		31.52	
	CLIP-Adapter	33.26	36.48	42.06
	Tip-Adapter	70.05	69.57	77.68
	Tip-Adapter-F	71.74	74.51	80.43
	Standard LP	71.07	76.30	80.34
	LP++	67.79	72.40	78.32
	CoOp	68.66	75.00	79.63
	CoCoOp	58.58	66.58	72.16
	KgCoOp	68.77	69.50	73.58
	ProGrad	69.13	70.99	75.11
	BiomedCoOp	71.19	74.78	79.05
	BiomedHELIX (Ours)	74.20	82.77	87.10
LC25000	BiomedCLIP		50.03	
	BiomedCLIP + Ensemble		61.84	
	CLIP-Adapter	52.91	56.33	57.56
	Tip-Adapter	83.32	87.25	89.17
	Tip-Adapter-F	79.57	90.41	92.35
	Standard LP	85.30	90.24	92.77
	LP++	82.61	89.14	92.58
	CoOp	84.66	87.50	92.19
	CoCoOp	77.44	85.57	87.38
	KgCoOp	82.10	84.63	86.79
	ProGrad	84.72	87.86	90.70
	BiomedCoOp	85.60	88.77	92.68
	BiomedHELIX (Ours)	88.43	92.17	95.27
RETINA	BiomedCLIP		26.26	
	BiomedCLIP + Ensemble		39.27	
	CLIP-Adapter	26.07	25.84	26.05
	Tip-Adapter	43.42	48.08	54.23
	Tip-Adapter-F	47.37	56.07	62.85
	Standard LP	51.31	53.94	62.27
	LP++	46.95	53.44	60.62
	CoOp	42.22	51.87	59.38
	CoCoOp	39.75	48.45	53.91
	KgCoOp	42.61	49.97	51.18
	ProGrad	43.09	52.26	50.47
	BiomedCoOp	45.58	56.47	61.28
	BiomedHELIX (Ours)	49.77	57.10	70.23

Table 7 (continued): Per-dataset performance comparison.

Dataset	Method	$K = 4$	$K = 8$	$K = 16$
KneeXray	BiomedCLIP		29.53	
	BiomedCLIP + Ensemble		39.37	
	CLIP-Adapter	28.96	28.80	29.08
	Tip-Adapter	24.19	25.76	33.17
	Tip-Adapter-F	26.59	26.46	27.67
	Standard LP	27.83	22.20	23.97
	LP++	28.92	23.75	26.38
	CoOp	23.85	26.23	28.48
	CoCoOp	30.66	21.78	24.86
	KgCoOp	22.44	23.37	24.80
	ProGrad	23.95	24.78	26.27
	BiomedCoOp	35.91	37.70	39.69
	BiomedHELIX (Ours)	24.83	32.00	36.90
OCTMNIST	BiomedCLIP		30.00	
	BiomedCLIP + Ensemble		47.40	
	CLIP-Adapter	49.96	49.50	52.73
	Tip-Adapter	38.10	53.93	53.33
	Tip-Adapter-F	55.20	65.00	72.50
	Standard LP	61.00	65.85	69.40
	LP++	59.02	63.69	68.35
	CoOp	53.37	63.67	65.47
	CoCoOp	48.57	55.40	60.67
	KgCoOp	52.97	61.03	62.80
	ProGrad	55.07	62.17	63.33
	BiomedCoOp	54.73	58.87	66.93
	BiomedHELIX (Ours)	61.27	77.07	83.77
Average	BiomedCLIP		42.05	
	BiomedCLIP + Ensemble		52.27	
	CLIP-Adapter	44.36	45.42	46.69
	Tip-Adapter	57.33	61.98	67.15
	Tip-Adapter-F	61.23	65.91	70.91
	Standard LP	61.00	65.85	69.40
	LP++	59.02	63.69	68.35
	CoOp	59.75	65.84	69.62
	CoCoOp	54.69	61.08	65.09
	KgCoOp	58.59	63.65	64.88
	ProGrad	60.42	65.61	67.13
	BiomedCoOp	63.95	68.32	72.42
	BiomedHELIX (Ours)	63.94	71.17	77.26