BiomedCoOp: Learning to Prompt for Biomedical Vision-Language Models

Supplementary Material

1. Detailed Dataset Overview

Table S7 provides a summary of the 11 datasets used for our proposed BiomedCoOp, covering 9 biomedical imaging modalities, such as CT, MRI, X-ray, ultrasound, and others, and 10 different organs. Each dataset is described in terms of its imaging modality, target organ(s), number of classes, and dataset splits (train/validation/test). The datasets span diverse clinical cases, including kidney cysts in CT, various skin lesions in dermatoscopy, and different stages of knee osteoarthritis in X-ray. These datasets capture a wide range of disease classes and imaging types, offering a rich and representative benchmark for biomedical image classification tasks. Instead of using full training splits, we employ random few-shot seeds to ensure efficient and representative learning from limited data. Additionally, the examples for each class are proportionally distributed across the splits, ensuring balanced representation, which enhances model evaluation on clinically relevant data and strengthens BiomedCoOp's robustness across diverse tasks.

2. Additional Few-shot Results

Figure S1 demonstrates the performance variations of BiomedCoOp and the baseline models under different few-shot configurations (K = 1,2,4,8,16). It underscores BiomedCoOp's robustness in adapting to limited data. On the other hand, we provide the detailed few-shot evaluation results for each dataset in Table S8. Overall, BiomedCoOp is on par and regularly outperforms SOTA parameter-efficient techniques across diverse datasets.

3. Learnable Context Interpretability

In this experiment, we aim to investigate the closest words to each of the four learned context tokens in various biomedical datasets, examining how these nearest words align with visual or anatomical characteristics in the images. This could offer some intuitive interpretation for the learned context, which is more abstract than a typical phrase like "`A photo of [CLASS]`". As shown in Table S9, the nearest words to each learned context token are listed, along with their corresponding Euclidean distances in the embedding space (in parentheses). It's particularly intriguing that some learned embeddings capture relevant descriptors, such as "`endoscopy`" for Kvasir, "`mri`" for BTMRI, or "`receptive`" for RETINA, reflecting contextual understanding of these biomedical imaging types.

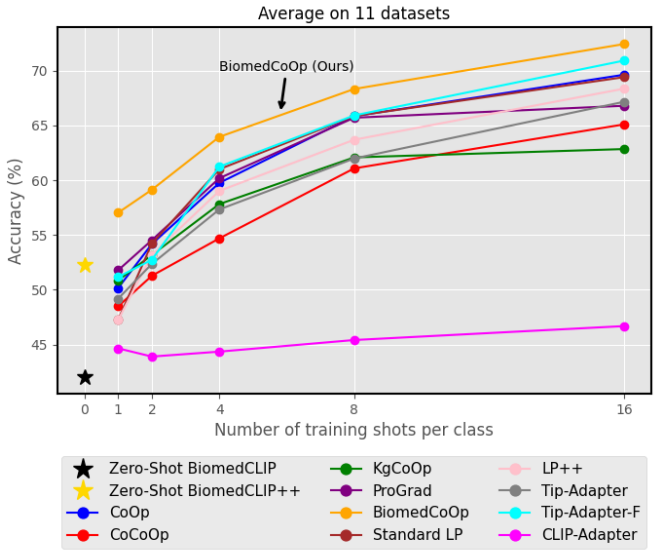

Figure S1. Average classification accuracy (%) of various few-shot adaptation methods across different numbers of training shots per class.

4. Effect of Context Length

As shown in Table S1, increasing the context length tends to reduce performance on both base and novel classes. A shorter context length, such as 4, achieves a better balance between base and novel accuracy, resulting in a higher harmonic mean (HM) score. As the context length increases to 16, 32, and 64, the accuracy on novel classes declines more rapidly than on base classes, leading to a sharp reduction in the harmonic mean. This pattern suggests that longer context lengths could diminish the model's ability to generalize effectively across both base and novel classes.

Context Length	Base Acc.	Novel Acc.	HM
4	76.11	73.22	74.64
16	74.93	67.98	71.29
32	72.34	62.73	67.19
64	71.50	58.99	64.65

Table S1. Effect of the context vector length on classification accuracy (%) in Base-to-Novel generalization.

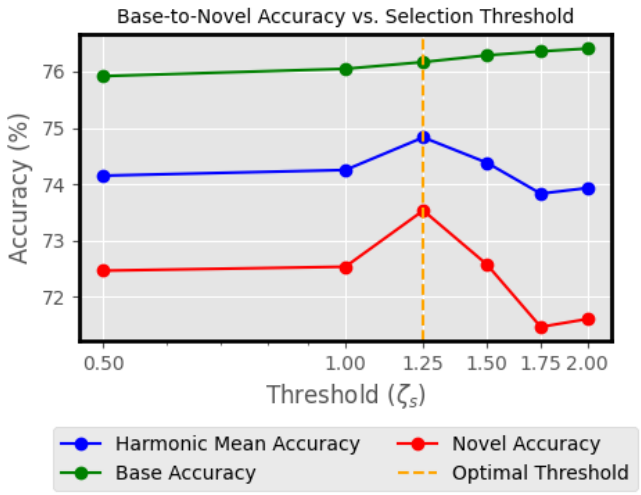

Figure S2. Effect of selection threshold (ζ_s) on Base-to-Novel Generalization

5. Effect of Prompt Selection Threshold

In the *Knowledge Distillation with Selective Prompting* (KDSP) component of our proposed method, we used a statistics-based prompt selection strategy. Figure S2 illustrates the impact of increasing the selection threshold (ζ_s) for the absolute value of the modified z-score to allow more prompts generated by the LLM to be used. This can lead to an overfitting effect, where the model becomes highly specialized in distinguishing base classes at the expense of its ability to generalize to novel classes. As a result, while the base class accuracy remains high or slightly increases with an increasing threshold, the accuracy for novel classes declines, as shown by the peak in novel accuracy at the optimal threshold (ζ_s=1.25). Beyond this point, higher thresholds reduce generalization capability, leading to decreased harmonic mean accuracy.

6. Selective Prompting for SCCM

We didn't perform prompt selection in the *Semantic Consistency by Contextual Mapping* (SCCM) component of our proposed BiomedCoOp framework. To verify the effect of prompt selection for SCCM, we compare the model performance with and without prompt selection. Table S5 shows that excluding outliers during contextual mapping marginally improves accuracy on base classes (76.26% to 76.39%), but reduces accuracy on novel classes (73.92% to 72.59%), resulting in a lower harmonic mean (74.44% to 73.92%). This suggests that outlier exclusion when applied to the SCCM component causes the prompts to overfit to base classes, reducing their flexibility and hindering generalization to novel classes. Thus, keeping all prompt samples in the mapping process (i.e., SCCM) helps maintain broader generalization, balancing performance across both base and novel classes.

7. Additional Comparisons with Other Recent Methods

We compared our method with two more recent SOTA CLIP adaptation methods (XCoOp [3] and DCPL [6]) on all datasets in Tables S2 and S3. We also compare with zero-shot methods (in blue). All methods were tuned to their optimal settings for each dataset, and 16-shot setting is used for all in Table S3. Specifically, we used an alternate version of DCPL, denoted DCPL*, which uses BiomedCLIP as the LSDM with deep multimodal prompting, whereas XCoOp directly utilizes the BiomedCLIP backbone. Our results demonstrate that our method consistently outperforms these approaches. On the other hand, average ensemble of LLM prompts improves zero-shot classification, but is suboptimal to prompt selection (Table S2).

Table S2. Avg. accuracy (%) comparison of additional SOTA methods in few-shot learning. DCPL* utilizes BiomedCLIP as the LSDM for domain knowledge while XCoOp directly utilizes the BiomedCLIP backbone. Methods in blue use zero-shot setting.

Method	$K=1$	$K=2$	$K=4$	$K=8$	$K=16$
BiomedCLIP	42.05				
BiomedCLIP + Ensemble	52.27 (+10.22 from BiomedCLIP)				
BiomedCLIP + Selection	53.72 (+11.67 from BiomedCLIP)				
DCPL*	$45.65_{8.86}$	$51.65_{8.79}$	$56.62_{7.51}$	$62.85_{8.40}$	$68.79_{4.80}$
XCoOp	$52.50_{5.91}$	$55.39_{5.74}$	$60.87_{4.18}$	$66.37_{3.44}$	$71.04_{1.95}$
BiomedCoOp (Ours)	$57.03_{2.80}$	$59.13_{3.64}$	$63.95_{2.42}$	$68.32_{2.65}$	$72.42_{1.69}$

Table S3. Base-to-novel generalization for recent SOTA methods. DCPL* utilizes BiomedCLIP as the LSDM for domain knowledge while XCoOp directly utilizes the BiomedCLIP backbone.

Method (K=16)	**Base Acc.**	**Novel Acc.**	**HM**
XCoOp	74.62%	63.19%	68.43%
DCPL*	68.70%	40.35%	50.84%
BiomedCoOp (Ours)	**76.26%**	**73.92%**	**75.07%**

8. Effect of LLM used

Table S4 presents an ablation study using 50 prompts from three recent LLMs across three datasets under the 4-shot evaluation. The results show that our model is robust to different text distributions, even with smaller LLMs like Gemma-2-2b, highlighting the benefit of our selection strategy in learning vigorous representations.

Table S4. Effect of different LLM choices on 4-shot accuracy (%)

Method	BTMRI	COVID-QU-Ex	CTKidney
LLaMA-3-8b	$76.61_{3.53}$	$73.20_{1.84}$	$\mathbf{67.81_{0.39}}$
Gemma-2-2b	$76.69_{3.46}$	$72.46_{2.46}$	$66.03_{0.96}$
GPT-4	$\mathbf{77.23_{3.90}}$	$\mathbf{73.28_{2.30}}$	$66.50_{1.92}$

Component	**Base Acc.**	**Novel Acc.**	**HM**
SCCM without SPOE	76.26	73.92	75.07
SCCM with SPOE	76.39	72.59	74.44

Table S5. Effect of excluding outliers in the SCCM block on classification accuracy (%) in Base-to-Novel generalization. SPOE = Selective Prompting via Outlier Exclusion. HM = harmonic mean.

9. Additional Hyperparameters

Table S6 outlines the selected hyperparameters (λ_1, λ_2, and ζ_s) used across various datasets for BiomedCoOp's few-shot and base-to-novel benchmarks. These parameters were optimized to balance classification accuracy and model adaptability, with λ_1 and λ_2 controlling the weight of the consistency and distillation losses, respectively, and ζ_s setting the selection threshold for prompt refinement. The selection threshold ζ_s remains consistently in the range [1.25, 1.5] across most datasets, indicating a stable value for effective prompt selection.

Dataset	Benchmark	λ_1	λ_2	ζ_s
BTMRI	Few-shot	0.5	0.25	1.5
	Base-to-Novel	0.5	0.5	1.25
BUSI	Few-shot	0.75	0.75	1.5
	Base-to-Novel	-	-	-
COVID-QU-Ex	Few-shot	0.5	2.0	1.5
	Base-to-Novel	20.0	1.0	1.25
CTKIDNEY	Few-shot	1.0	0.5	1.5
	Base-to-Novel	10.0	0.25	1.25
DermaMNIST	Few-shot	5.0	20.0	1.5
	Base-to-Novel	2.0	0.5	1.5
Kvasir	Few-shot	0.75	0.75	1.5
	Base-to-Novel	1.0	1.0	1.25
CHMNIST	Few-shot	0.25	0.25	1.5
	Base-to-Novel	10.0	1.0	1.5
LC25000	Few-shot	0.5	0.5	1.5
	Base-to-Novel	0.25	0.75	1.25
RETINA	Few-shot	0.25	0.25	1.5
	Base-to-Novel	5.0	1.0	2.0
KneeXray	Few-shot	5.0	20.0	1.75
	Base-to-Novel	0.25	3.0	1.25
OCTMNIST	Few-shot	1.0	0.75	1.5
	Base-to-Novel	0.75	0.5	1.5

Table S6. Hyperparameter values for λ_1, λ_2, and ζ_s across different datasets and benchmarks.

10. LLM Prompts Used

We include here one text prompt generated from GPT-4 for each class across all the datasets:

"The image of a normal brain on MRI shows a clear differentiation between different brain regions with no disruptions."

"Central necrosis and surrounding edema in glioma tumor on MRI scan."

"Meningioma tumor on MRI displaying a dural tail sign and homogeneous enhancement."

"Pituitary tumors often cause sellar expansion and may invade adjacent structures."

"A routine ultrasound showing a hypoechoic, well-defined nodule, indicating a benign breast tumor."

"An ultrasound revealing microcalcifications within the mass, indicating a malignant breast tumor."

"A grayscale ultrasound highlighting well-defined ducts and lobules, characteristic of a normal breast ultrasound scan."

"An X-ray scan showing bilateral airspace consolidation, typical of covid lungs."

"A chest X-ray image with reticular and nodular opacities, indicative of lung opacity lungs."

"An X-ray revealing no signs of consolidation or effusion, suggesting normal lungs."

“An X-ray image revealing multifocal ground-glass and consolidative opacities, indicative of viral pneumonia lungs.”

“A CT image showing a lesion with uniform density and no internal irregularities, indicative of a cyst kidney.”

“A CT scan showing a calcified structure with acoustic shadowing, consistent with a kidney stone.”

“A CT scan showing a lesion with poorly defined margins, consistent with a kidney tumor.”

“A CT image revealing no signs of renal atrophy or cortical thinning, suggesting a normal kidney.”

“Actinic keratosis lesions may become thicker and more pronounced over time without treatment.”

“BCC lesions may bleed with minor trauma, such as shaving, due to their friable nature.”

“Cryotherapy, using liquid nitrogen, is a common treatment for seborrheic keratosis, causing the lesions to blister and fall off.”

“Dermatofibromas can be multiple in patients with systemic lupus erythematosus or other autoimmune conditions.”

“A clinical image with a lesion that has changed in size or texture, indicative of melanoma.”

“Melanocytic nevi can become darker and larger during pregnancy due to hormonal changes and increased melanin production.”

“The diagnosis of vascular lesions often requires a combination of clinical examination and sometimes imaging studies.”

“Dyed lifted polyps can exhibit various morphological features, including lobulated, sessile, or pedunculated appearances.”

“Endoscopic images of dyed resection margins often show a bright, distinct color outlining the area of resection, contrasting with the surrounding mucosa.”

“In severe cases, esophagitis may lead to strictures or narrowing of the esophageal lumen, visible during endoscopy.”

“Endoscopic images of the normal cecum show a well-defined junction with the ascending colon, without any transitional abnormalities.”

“Endoscopic examination of the normal pylorus shows a lack of any masses, polyps, or other abnormal growths.”

“The Z line in a normal endoscopy appears intact and well-defined, with no evidence of structural compromise.”

“Polyps can be classified based on their appearance and histological features, including adenomatous polyps, hyperplastic polyps, or inflammatory polyps.”

“Ulcerative colitis can be associated with extra-intestinal manifestations, including dermatological, joint, ocular, or hepatobiliary complications.”

Modality	Organ(s)	Name	Classes	# train/val/test
Computerized Tomography	Kidney	CTKidney [20]	Kidney Cyst, Kidney Stone, Kidney Tumor, Normal Kidney	6221/2487/3738
Dermatoscopy	Skin	DermaMNIST [10, 40]	Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis, Dermatofibroma, Melanocytic nevus, Melanoma, Vascular Lesion	7007/1003/2005
Endoscopy	Colon	Kvasir [35]	Dyed Lifted Polyps, Normal Cecum, Esophagitis, Dyed Resection Margins, Normal Pylorus, Normal Z Line, Polyps, Ulcerative Colitis	2000/800/1200
Fundus Photography	Retina	RETINA [31, 36]	Cataract, Diabetic Retinopathy, Glaucoma, Normal Retina	2108/841/1268
Histopathology	Lung Colon	LC25000 [5]	Colon Adenocarcinoma, Colon Benign Tissue, Lung Adenocarcinoma, Lung Benign Tissue, Lung Squamous Cell Carcinoma	12500/5000//7500
	Colorectal	CHMNIST [22]	Adipose Tissue, Complex Stroma, Debris, Empty Background, Immune Cells, Normal Mucosal Glands, Simple Stroma, Tumor Epithelium	2496/1000/1504
Magnetic Resonance Imaging	Brain	BTMRI [34]	Glioma Tumor, Meningioma Tumor, Normal Brain, Pituitary Tumor	2854/1141/1717
Optical Coherence Tomography	Retina	OCTMNIST [23]	Choroidal Neovascularization, Drusen, Diabetic Macular Edema, Normal	97477/10832/1000
Ultrasound	Breast	BUSI [2]	Benign Tumors, Malignant Tumors, Normal Scans	389/155/236
X-Ray	Chest	COVID-QU-Ex [39]	COVID-19, Lung Opacity, Normal Lungs, Viral Pneumonia	10582/4232/6351
	Knee	KneeXray [7]	No, Doubtful, Minimal, Moderate, and Severe Osteoarthritis	5778/826/1656

Table S7. An overview of the 11 datasets used spanning 9 biomedical imaging modalities and 10 different organs.

Dataset	Method	$K = 1$	$K = 2$	$K = 4$	$K = 8$	$K = 16$
BTMRI	BiomedCLIP			56.79		
	BiomedCLIP + Ensemble			61.04		
	CLIP-Adapter	$56.80_{\pm 0.48}$	$57.13_{\pm 0.88}$	$56.80_{\pm 0.48}$	$57.15_{\pm 0.91}$	$60.16_{\pm 0.32}$
	Tip-Adapter	$66.66_{\pm 4.37}$	$67.77_{\pm 2.74}$	$76.37_{\pm 1.69}$	$73.75_{\pm 3.15}$	$78.97_{\pm 1.25}$
	Tip-Adapter-F	$59.60_{\pm 2.28}$	$61.94_{\pm 6.74}$	$77.90_{\pm 1.71}$	$79.18_{\pm 1.80}$	$82.27_{\pm 2.33}$
	Standard LP	$62.24_{\pm 5.03}$	$72.45_{\pm 5.27}$	$75.98_{\pm 1.94}$	$77.63_{\pm 3.45}$	$81.24_{\pm 2.56}$
	LP++	$64.72_{\pm 6.16}$	$71.69_{\pm 5.88}$	$75.48_{\pm 1.41}$	$77.11_{\pm 1.28}$	$81.61_{\pm 1.31}$
	CoOp	$63.82_{\pm 3.94}$	$68.82_{\pm 5.15}$	$74.68_{\pm 2.99}$	$79.27_{\pm 1.9}$	$82.37_{\pm 1.89}$
	CoCoOp	$59.47_{\pm 0.78}$	$64.14_{\pm 0.64}$	$67.83_{\pm 4.8}$	$71.69_{\pm 4.4}$	$78.45_{\pm 1.83}$
	KgCoOp	$63.33_{\pm 3.66}$	$70.16_{\pm 5.47}$	$75.4_{\pm 2.45}$	$79.79_{\pm 0.99}$	$81.07_{\pm 0.33}$
	ProGrad	$66.92_{\pm 2.10}$	$71.46_{\pm 3.46}$	$76.24_{\pm 5.07}$	$78.82_{\pm 1.77}$	$82.84_{\pm 1.02}$
	BiomedCoOp (Ours)	$65.08_{\pm 1.81}$	$70.57_{\pm 4.31}$	$77.23_{\pm 3.9}$	$78.55_{\pm 2.19}$	$83.3_{\pm 1.34}$
BUSI	BiomedCLIP			59.75		
	BiomedCLIP + Ensemble			59.75		
	CLIP-Adapter	$61.44_{\pm 0.78}$	$61.01_{\pm 1.03}$	$61.72_{\pm 0.81}$	$61.86_{\pm 1.41}$	$63.55_{\pm 2.17}$
	Tip-Adapter	$62.71_{\pm 2.56}$	$61.44_{\pm 2.44}$	$59.03_{\pm 1.13}$	$55.93_{\pm 11.37}$	$68.78_{\pm 5.54}$
	Tip-Adapter-F	$61.86_{\pm 2.17}$	$56.35_{\pm 7.25}$	$64.54_{\pm 7.01}$	$68.50_{\pm 2.26}$	$71.89_{\pm 1.25}$
	Standard LP	$51.41_{\pm 10.78}$	$47.88_{\pm 6.44}$	$53.38_{\pm 7.12}$	$65.53_{\pm 6.34}$	$68.78_{\pm 1.80}$
	LP++	$51.12_{\pm 4.95}$	$55.50_{\pm 2.38}$	$60.31_{\pm 3.42}$	$66.10_{\pm 2.34}$	$70.05_{\pm 1.58}$
	CoOp	$48.73_{\pm 3.3}$	$53.53_{\pm 2.8}$	$60.17_{\pm 3.65}$	$64.69_{\pm 6.4}$	$69.49_{\pm 3.3}$
	CoCoOp	$52.26_{\pm 3.73}$	$49.15_{\pm 2.77}$	$59.75_{\pm 1.83}$	$65.82_{\pm 3.83}$	$70.2_{\pm 1.22}$
	KgCoOp	$53.39_{\pm 7.25}$	$55.51_{\pm 3.30}$	$62.01_{\pm 4.38}$	$67.37_{\pm 2.42}$	$70.62_{\pm 2.11}$
	ProGrad	$46.33_{\pm 4.23}$	$49.15_{\pm 7.32}$	$62.29_{\pm 7.49}$	$64.83_{\pm 4.20}$	$71.47_{\pm 2.69}$
	BiomedCoOp (Ours)	$50.71_{\pm 1.74}$	$50.71_{\pm 7.34}$	$59.32_{\pm 1.04}$	$63.27_{\pm 4.61}$	$70.34_{\pm 2.27}$
COVID-QU-Ex	BiomedCLIP			43.8		
	BiomedCLIP + Ensemble			66.86		
	CLIP-Adapter	$50.42_{\pm 1.55}$	$43.04_{\pm 1.16}$	$46.28_{\pm 3.30}$	$48.68_{\pm 1.13}$	$49.55_{\pm 1.35}$
	Tip-Adapter	$62.13_{\pm 7.82}$	$58.72_{\pm 5.19}$	$63.84_{\pm 10.41}$	$66.77_{\pm 5.64}$	$73.05_{\pm 1.04}$
	Tip-Adapter-F	$54.89_{\pm 17.51}$	$54.01_{\pm 7.87}$	$69.97_{\pm 4.13}$	$69.89_{\pm 4.08}$	$76.07_{\pm 3.22}$
	Standard LP	$49.91_{\pm 10.98}$	$48.06_{\pm 16.94}$	$60.55_{\pm 13.60}$	$68.29_{\pm 6.12}$	$71.98_{\pm 1.88}$
	LP++	$46.41_{\pm 10.75}$	$56.42_{\pm 15.04}$	$62.32_{\pm 9.54}$	$66.19_{\pm 8.40}$	$72.79_{\pm 1.17}$
	CoOp	$58.82_{\pm 14.51}$	$58.37_{\pm 8.14}$	$67.03_{\pm 6.58}$	$74.66_{\pm 0.29}$	$76.37_{\pm 1.39}$
	CoCoOp	$69.36_{\pm 2.79}$	$68.8_{\pm 2.65}$	$63.7_{\pm 10.27}$	$69.36_{\pm 3.28}$	$74.52_{\pm 0.72}$
	KgCoOp	$61.68_{\pm 9.84}$	$54.68_{\pm 12.19}$	$65.91_{\pm 8.61}$	$74.86_{\pm 0.28}$	$75.65_{\pm 0.88}$
	ProGrad	$60.42_{\pm 11.74}$	$64.22_{\pm 6.44}$	$68.56_{\pm 3.2}$	$74.65_{\pm 1.09}$	$74.93_{\pm 1.07}$
	BiomedCoOp (Ours)	$72.64_{\pm 2.41}$	$71.53_{\pm 1.5}$	$73.28_{\pm 2.30}$	$76.26_{\pm 0.38}$	$78.72_{\pm 0.23}$
CTKIDNEY	BiomedCLIP			42.43		
	BiomedCLIP + Ensemble			56.82		
	CLIP-Adapter	$47.17_{\pm 3.74}$	$41.94_{\pm 2.15}$	$42.19_{\pm 2.27}$	$44.64_{\pm 0.90}$	$47.28_{\pm 1.41}$
	Tip-Adapter	$45.85_{\pm 5.41}$	$51.65_{\pm 7.87}$	$55.33_{\pm 4.10}$	$69.89_{\pm 8.74}$	$73.38_{\pm 7.77}$
	Tip-Adapter-F	$46.68_{\pm 6.70}$	$58.99_{\pm 8.54}$	$60.18_{\pm 10.73}$	$75.24_{\pm 6.89}$	$82.07_{\pm 3.29}$
	Standard LP	$43.82_{\pm 6.43}$	$59.35_{\pm 6.49}$	$69.54_{\pm 7.67}$	$78.89_{\pm 7.37}$	$82.50_{\pm 5.22}$
	LP++	$57.70_{\pm 2.85}$	$61.57_{\pm 3.38}$	$65.73_{\pm 9.15}$	$77.06_{\pm 7.96}$	$79.07_{\pm 7.67}$
	CoOp	$54.51_{\pm 8.74}$	$60.57_{\pm 2.26}$	$68.12_{\pm 2.11}$	$77.4_{\pm 3.87}$	$83.52_{\pm 1.8}$
	CoCoOp	$47.88_{\pm 7.72}$	$52.71_{\pm 9.71}$	$61.07_{\pm 1.33}$	$73.93_{\pm 1.5}$	$77.7_{\pm 2.65}$
	KgCoOp	$58.92_{\pm 1.28}$	$62.81_{\pm 3.38}$	$68.68_{\pm 5.54}$	$77.43_{\pm 4.2}$	$77.67_{\pm 3.12}$
	ProGrad	$54.65_{\pm 8.97}$	$64.66_{\pm 5.31}$	$67.90_{\pm 2.02}$	$78.23_{\pm 4.74}$	$81.13_{\pm 2.28}$
	BiomedCoOp (Ours)	$56.13_{\pm 4.19}$	$64.21_{\pm 5.57}$	$66.5_{\pm 1.92}$	$77.16_{\pm 3.98}$	$83.20_{\pm 2.37}$
DermaMNIST	BiomedCLIP			38.75		
	BiomedCLIP + Ensemble			53.62		
	CLIP-Adapter	$35.96_{\pm 6.70}$	$36.01_{\pm 6.63}$	$34.97_{\pm 4.17}$	$34.28_{\pm 6.55}$	$29.02_{\pm 3.80}$
	Tip-Adapter	$37.52_{\pm 2.12}$	$40.98_{\pm 13.52}$	$47.31_{\pm 6.23}$	$61.67_{\pm 5.79}$	$62.67_{\pm 0.97}$
	Tip-Adapter-F	$37.34_{\pm 15.72}$	$38.52_{\pm 4.39}$	$50.44_{\pm 5.30}$	$43.87_{\pm 2.18}$	$53.86_{\pm 4.99}$
	Standard LP	$30.67_{\pm 13.12}$	$38.13_{\pm 10.28}$	$49.77_{\pm 8.34}$	$51.02_{\pm 2.99}$	$55.34_{\pm 3.56}$
	LP++	$26.93_{\pm 3.93}$	$26.16_{\pm 11.70}$	$36.29_{\pm 9.19}$	$45.78_{\pm 2.74}$	$50.98_{\pm 2.14}$
	CoOp	$25.88_{\pm 9.07}$	$38.92_{\pm 6.01}$	$43.71_{\pm 6.27}$	$46.8_{\pm 6.80}$	$51.07_{\pm 2.56}$
	CoCoOp	$24.51_{\pm 4.22}$	$24.96_{\pm 0.76}$	$25.29_{\pm 5.61}$	$40.42_{\pm 2.44}$	$40.97_{\pm 6.50}$
	KgCoOp	$27.1_{\pm 10.81}$	$30.28_{\pm 4.45}$	$35.35_{\pm 8.07}$	$38.79_{\pm 4.85}$	$36.59_{\pm 2.32}$
	ProGrad	$33.98_{\pm 10.76}$	$37.66_{\pm 6.74}$	$43.69_{\pm 10.96}$	$51.07_{\pm 2.47}$	$46.33_{\pm 5.13}$
	BiomedCoOp (Ours)	$58.64_{\pm 4.71}$	$57.17_{\pm 1.28}$	$60.07_{\pm 1.81}$	$61.98_{\pm 0.77}$	$62.59_{\pm 1.83}$

Table S8. Per-dataset performance comparison of BiomedCoOp with various methods in few-shot setting in terms of classification accuracy (%).

Dataset	Method	$K = 1$	$K = 2$	$K = 4$	$K = 8$	$K = 16$
Kvasir	BiomedCLIP			54.58		
	BiomedCLIP + Ensemble			57.5		
	CLIP-Adapter	$54.83_{\pm 0.48}$	$54.83_{\pm 0.48}$	$54.83_{\pm 0.48}$	$56.08_{\pm 0.86}$	$56.50_{\pm 1.00}$
	Tip-Adapter	$56.72_{\pm 3.42}$	$60.94_{\pm 5.30}$	$69.61_{\pm 2.06}$	$69.13_{\pm 1.44}$	$74.22_{\pm 1.51}$
	Tip-Adapter-F	$59.19_{\pm 0.89}$	$64.22_{\pm 3.24}$	$69.94_{\pm 2.28}$	$75.86_{\pm 1.00}$	$78.00_{\pm 1.06}$
	Standard LP	$54.30_{\pm 2.04}$	$62.00_{\pm 0.81}$	$72.38_{\pm 2.65}$	$78.88_{\pm 0.73}$	$79.00_{\pm 0.81}$
	LP++	$58.27_{\pm 3.95}$	$60.47_{\pm 3.24}$	$69.36_{\pm 0.84}$	$72.52_{\pm 2.89}$	$75.41_{\pm 1.21}$
	CoOp	$58.2_{\pm 1.64}$	$64.86_{\pm 1.4}$	$70.78_{\pm 0.31}$	$77.14_{\pm 1.25}$	$77.88_{\pm 0.12}$
	CoCoOp	$59.45_{\pm 3.25}$	$65.5_{\pm 3.41}$	$68.94_{\pm 1.29}$	$72.92_{\pm 1.46}$	$75.22_{\pm 2.04}$
	KgCoOp	$61.67_{\pm 2.16}$	$65.67_{\pm 1.94}$	$68.28_{\pm 0.35}$	$72.05_{\pm 1.8}$	$72.95_{\pm 1.31}$
	ProGrad	$60.78_{\pm 0.24}$	$64.70_{\pm 0.53}$	$70.00_{\pm 0.24}$	$76.03_{\pm 1.50}$	$75.88_{\pm 0.95}$
	BiomedCoOp (Ours)	$62.17_{\pm 1.95}$	$67.25_{\pm 2.59}$	$74.08_{\pm 1.10}$	$77.72_{\pm 0.52}$	$78.89_{\pm 1.21}$
CHMNIST	BiomedCLIP			30.65		
	BiomedCLIP + Ensemble			31.52		
	CLIP-Adapter	$31.27_{\pm 0.69}$	$31.67_{\pm 0.88}$	$33.26_{\pm 0.39}$	$36.48_{\pm 1.32}$	$42.06_{\pm 2.40}$
	Tip-Adapter	$46.14_{\pm 9.62}$	$63.32_{\pm 2.58}$	$70.05_{\pm 1.11}$	$69.57_{\pm 1.63}$	$77.68_{\pm 1.42}$
	Tip-Adapter-F	$52.81_{\pm 3.10}$	$58.90_{\pm 4.95}$	$71.74_{\pm 2.72}$	$74.51_{\pm 2.43}$	$80.43_{\pm 2.85}$
	Standard LP	$58.44_{\pm 2.02}$	$64.42_{\pm 3.81}$	$71.07_{\pm 2.23}$	$76.30_{\pm 3.22}$	$80.34_{\pm 1.83}$
	LP++	$57.18_{\pm 6.46}$	$60.61_{\pm 1.26}$	$67.79_{\pm 6.97}$	$72.40_{\pm 0.85}$	$78.32_{\pm 1.48}$
	CoOp	$57.34_{\pm 4.2}$	$59.68_{\pm 1.12}$	$68.66_{\pm 2.14}$	$75.00_{\pm 0.82}$	$79.63_{\pm 1.26}$
	CoCoOp	$49.07_{\pm 4.41}$	$50.82_{\pm 3.41}$	$58.58_{\pm 2.15}$	$66.58_{\pm 1.14}$	$72.16_{\pm 0.52}$
	KgCoOp	$59.02_{\pm 4.1}$	$60.06_{\pm 1.12}$	$68.77_{\pm 1.02}$	$69.50_{\pm 0.07}$	$73.58_{\pm 1.19}$
	ProGrad	$60.15_{\pm 5.76}$	$59.60_{\pm 1.53}$	$69.13_{\pm 1.39}$	$70.99_{\pm 0.36}$	$75.11_{\pm 1.50}$
	BiomedCoOp (Ours)	$59.82_{\pm 2.43}$	$59.79_{\pm 1.36}$	$71.19_{\pm 1.74}$	$74.78_{\pm 1.19}$	$79.05_{\pm 2.24}$
LC25000	BiomedCLIP			50.03		
	BiomedCLIP + Ensemble			61.84		
	CLIP-Adapter	$54.83_{\pm 0.36}$	$53.47_{\pm 2.95}$	$52.91_{\pm 1.70}$	$56.33_{\pm 0.45}$	$57.56_{\pm 1.13}$
	Tip-Adapter	$75.37_{\pm 4.02}$	$72.73_{\pm 8.09}$	$83.32_{\pm 3.95}$	$87.25_{\pm 1.75}$	$89.17_{\pm 0.41}$
	Tip-Adapter-F	$74.21_{\pm 4.35}$	$71.82_{\pm 7.31}$	$79.57_{\pm 10.02}$	$90.41_{\pm 2.43}$	$92.35_{\pm 1.08}$
	Standard LP	$74.50_{\pm 2.61}$	$78.40_{\pm 7.36}$	$85.30_{\pm 3.56}$	$90.24_{\pm 0.41}$	$92.77_{\pm 1.17}$
	LP++	$63.05_{\pm 9.52}$	$71.42_{\pm 3.04}$	$82.61_{\pm 2.31}$	$89.14_{\pm 2.07}$	$92.58_{\pm 0.38}$
	CoOp	$71.90_{\pm 3.53}$	$76.55_{\pm 2.81}$	$84.66_{\pm 2.26}$	$87.50_{\pm 0.26}$	$92.19_{\pm 0.48}$
	CoCoOp	$63.66_{\pm 4.49}$	$71.76_{\pm 0.55}$	$77.44_{\pm 2.47}$	$85.57_{\pm 1.83}$	$87.38_{\pm 0.52}$
	KgCoOp	$71.80_{\pm 2.13}$	$75.18_{\pm 1.05}$	$82.10_{\pm 2.35}$	$84.63_{\pm 0.30}$	$86.79_{\pm 0.53}$
	ProGrad	$72.48_{\pm 3.22}$	$74.76_{\pm 1.40}$	$84.72_{\pm 2.85}$	$87.86_{\pm 0.70}$	$90.70_{\pm 0.66}$
	BiomedCoOp (Ours)	$77.56_{\pm 2.84}$	$77.74_{\pm 2.00}$	$85.60_{\pm 1.61}$	$88.77_{\pm 1.14}$	$92.68_{\pm 0.57}$
RETINA	BiomedCLIP			26.26		
	BiomedCLIP + Ensemble			39.27		
	CLIP-Adapter	$25.49_{\pm 0.46}$	$25.49_{\pm 0.46}$	$26.07_{\pm 0.46}$	$25.84_{\pm 0.87}$	$26.05_{\pm 0.43}$
	Tip-Adapter	$26.52_{\pm 0.42}$	$31.07_{\pm 3.84}$	$43.42_{\pm 7.04}$	$48.08_{\pm 7.40}$	$54.23_{\pm 5.13}$
	Tip-Adapter-F	$39.53_{\pm 10.83}$	$33.07_{\pm 5.63}$	$47.37_{\pm 6.70}$	$56.07_{\pm 2.57}$	$62.85_{\pm 1.10}$
	Standard LP	$39.35_{\pm 6.96}$	$46.03_{\pm 0.79}$	$51.31_{\pm 6.52}$	$53.94_{\pm 1.98}$	$62.27_{\pm 2.80}$
	LP++	$35.77_{\pm 5.75}$	$39.37_{\pm 7.35}$	$46.95_{\pm 10.07}$	$53.44_{\pm 1.95}$	$60.62_{\pm 1.46}$
	CoOp	$35.02_{\pm 1.40}$	$35.26_{\pm 3.34}$	$42.22_{\pm 3.09}$	$51.87_{\pm 1.78}$	$59.38_{\pm 0.87}$
	CoCoOp	$32.94_{\pm 0.75}$	$36.43_{\pm 4.05}$	$39.75_{\pm 3.99}$	$48.45_{\pm 1.39}$	$53.91_{\pm 1.52}$
	KgCoOp	$33.54_{\pm 2.77}$	$35.17_{\pm 2.48}$	$42.61_{\pm 3.16}$	$49.97_{\pm 2.24}$	$51.18_{\pm 1.66}$
	ProGrad	$33.49_{\pm 1.98}$	$36.49_{\pm 4.64}$	$43.09_{\pm 3.89}$	$52.26_{\pm 2.38}$	$50.47_{\pm 2.40}$
	BiomedCoOp (Ours)	$36.64_{\pm 3.34}$	$38.67_{\pm 1.79}$	$45.58_{\pm 5.03}$	$56.47_{\pm 1.37}$	$61.28_{\pm 1.06}$
KneeXray	BiomedCLIP			29.53		
	BiomedCLIP + Ensemble			39.37		
	CLIP-Adapter	$29.00_{\pm 0.17}$	$28.66_{\pm 0.45}$	$28.96_{\pm 0.46}$	$28.80_{\pm 0.20}$	$29.08_{\pm 0.32}$
	Tip-Adapter	$29.04_{\pm 0.67}$	$33.55_{\pm 5.96}$	$24.19_{\pm 4.23}$	$25.76_{\pm 3.35}$	$33.17_{\pm 7.59}$
	Tip-Adapter-F	$30.01_{\pm 0.50}$	$28.38_{\pm 2.18}$	$26.59_{\pm 5.70}$	$26.46_{\pm 2.20}$	$27.67_{\pm 3.21}$
	Standard LP	$26.02_{\pm 11.08}$	$26.57_{\pm 5.17}$	$27.83_{\pm 4.92}$	$22.20_{\pm 3.68}$	$23.97_{\pm 3.55}$
	LP++	$21.25_{\pm 8.60}$	$26.40_{\pm 3.26}$	$28.92_{\pm 4.97}$	$23.75_{\pm 2.50}$	$26.38_{\pm 3.39}$
	CoOp	$24.96_{\pm 9.41}$	$25.89_{\pm 5.06}$	$23.85_{\pm 4.25}$	$26.23_{\pm 4.01}$	$28.48_{\pm 1.84}$
	CoCoOp	$25.42_{\pm 6.38}$	$28.85_{\pm 8.24}$	$30.66_{\pm 4.49}$	$21.78_{\pm 8.29}$	$24.86_{\pm 4.15}$
	KgCoOp	$29.07_{\pm 3.31}$	$28.14_{\pm 4.53}$	$22.44_{\pm 2.88}$	$23.37_{\pm 3.35}$	$24.80_{\pm 0.47}$
	ProGrad	$30.09_{\pm 6.00}$	$23.83_{\pm 0.57}$	$23.95_{\pm 2.87}$	$24.78_{\pm 2.32}$	$26.27_{\pm 3.29}$
	BiomedCoOp (Ours)	$36.13_{\pm 1.75}$	$37.72_{\pm 0.54}$	$35.91_{\pm 0.54}$	$37.7_{\pm 1.00}$	$39.69_{\pm 1.75}$

Table S8 (continued): Per-dataset performance comparison of BiomedCoOp with various methods in few-shot setting in terms of classification accuracy (%).

Dataset	Method	$K = 1$	$K = 2$	$K = 4$	$K = 8$	$K = 16$
OCTMNIST	BiomedCLIP			30.00		
	BiomedCLIP + Ensemble			47.40		
	CLIP-Adapter	$44.00_{\pm 5.79}$	$49.73_{\pm 2.41}$	$49.96_{\pm 1.77}$	$49.50_{\pm 3.33}$	$52.73_{\pm 0.62}$
	Tip-Adapter	$32.36_{\pm 3.94}$	$33.8_{\pm 6.16}$	$38.10_{\pm 5.01}$	$53.93_{\pm 3.17}$	$53.33_{\pm 3.92}$
	Tip-Adapter-F	$46.66_{\pm 2.58}$	$53.93_{\pm 1.67}$	$55.20_{\pm 4.75}$	$65.00_{\pm 6.61}$	$72.50_{\pm 1.38}$
	Standard LP	$47.25_{\pm 12.64}$	$54.21_{\pm 8.23}$	$61.00_{\pm 7.07}$	$65.85_{\pm 9.01}$	$69.40_{\pm 3.68}$
	LP++	$47.24_{\pm 13.84}$	$53.18_{\pm 9.08}$	$59.02_{\pm 8.59}$	$63.69_{\pm 8.26}$	$68.35_{\pm 7.42}$
	CoOp	$52.63_{\pm 2.95}$	$53.57_{\pm 3.86}$	$53.37_{\pm 2.35}$	$63.67_{\pm 4.47}$	$65.47_{\pm 7.47}$
	CoCoOp	$49.33_{\pm 4.58}$	$50.93_{\pm 8.01}$	$48.57_{\pm 6.25}$	$55.40_{\pm 1.88}$	$60.67_{\pm 3.41}$
	KgCoOp	$50.63_{\pm 3.18}$	$50.53_{\pm 5.39}$	$52.97_{\pm 1.58}$	$61.03_{\pm 3.78}$	$62.80_{\pm 3.85}$
	ProGrad	$51.40_{\pm 3.05}$	$55.33_{\pm 3.38}$	$55.07_{\pm 1.22}$	$62.17_{\pm 6.01}$	$63.33_{\pm 6.15}$
	BiomedCoOp (Ours)	$51.83_{\pm 1.52}$	$55.03_{\pm 4.72}$	$54.73_{\pm 1.86}$	$58.87_{\pm 5.35}$	$66.93_{\pm 2.13}$
Average	BiomedCLIP			42.05		
	BiomedCLIP + Ensemble			52.27		
	CLIP-Adapter	$44.66_{\pm 2.97}$	$43.91_{\pm 2.48}$	$44.36_{\pm 1.94}$	$45.42_{\pm 2.38}$	$46.69_{\pm 1.71}$
	Tip-Adapter	$49.19_{\pm 4.84}$	$52.36_{\pm 6.57}$	$57.33_{\pm 5.07}$	$61.98_{\pm 5.76}$	$67.15_{\pm 4.25}$
	Tip-Adapter-F	$51.17_{\pm 8.33}$	$52.74_{\pm 5.88}$	$61.23_{\pm 6.22}$	$65.91_{\pm 3.64}$	$70.91_{\pm 2.65}$
	Standard LP	$47.25_{\pm 8.65}$	$54.21_{\pm 7.80}$	$61.00_{\pm 6.81}$	$65.85_{\pm 4.89}$	$69.40_{\pm 2.91}$
	LP++	$47.24_{\pm 7.68}$	$53.18_{\pm 7.29}$	$59.02_{\pm 6.93}$	$63.69_{\pm 4.68}$	$68.35_{\pm 3.59}$
	CoOp	$50.16_{\pm 6.93}$	$54.18_{\pm 4.31}$	$59.75_{\pm 3.72}$	$65.84_{\pm 3.66}$	$69.62_{\pm 2.83}$
	CoCoOp	$48.49_{\pm 4.39}$	$51.28_{\pm 5.06}$	$54.69_{\pm 4.79}$	$61.08_{\pm 3.49}$	$65.09_{\pm 2.87}$
	KgCoOp	$51.83_{\pm 5.53}$	$53.47_{\pm 5.07}$	$58.59_{\pm 4.50}$	$63.65_{\pm 2.73}$	$64.88_{\pm 1.95}$
	ProGrad	$51.88_{\pm 6.39}$	$54.71_{\pm 4.46}$	$60.42_{\pm 4.78}$	$65.61_{\pm 3.02}$	$67.13_{\pm 3.00}$
	BiomedCoOp (Ours)	$57.03_{\pm 2.80}$	$59.13_{\pm 3.64}$	$63.95_{\pm 2.42}$	$68.32_{\pm 2.65}$	$72.42_{\pm 1.62}$

Table S8 (continued): Per-dataset performance comparison of BiomedCoOp with various methods in few-shot setting in terms of classification accuracy (%).

Dataset	Context Token #1	Context Token #2	Context Token #3	Context Token #4
BTMRI	`mri` (2.4971)	`curcumin` (2.5835)	`of` (1.5667)	`a` (1.6353)
BUSI	`a` (2.5550)	`photo` (3.5649)	`of` (2.1298)	`b` (3.4897)
COVID-QU-Ex	`measured` (2.1999)	`image` (2.2856)	`of` (1.9166)	`a` (1.9205)
CTKIDNEY	`a` (2.1290)	`schem` (2.6564)	`right` (2.3790)	`a` (1.7574)
DermaMNIST	`dextrose` (2.8292)	`photo` (3.1084)	`ricin` (3.2378)	`autologous` (3.0297)
Kvasir	`endoscopy` (2.1880)	`scar` (2.4835)	`of` (2.2698)	`maintained` (2.4771)
CHMNIST	`a` (3.0301)	`original` (3.4248)	`composed` (2.2125)	`discern` (3.4506)
LC25000	`a` (1.5298)	`photo` (2.3540)	`of` (1.6363)	`a` (2.0292)
RETINA	`a` (1.5986)	`papill` (2.3636)	`of` (1.6976)	`receptive` (2.1135)
KneeXray	`a` (4.2063)	`calcification` (5.4999)	`osteoc` (2.8673)	`showed` (2.9774)
OCTMNIST	`localized` (2.1744)	`example` (3.6750)	`of` (1.8752)	`possible` (2.4803)

Table S9. The nearest words for each of the 4 context vectors learned by BiomedCoOp, with their distances to the corresponding context tokens shown in parentheses.