RIT
Rochester
Institute of
Technology

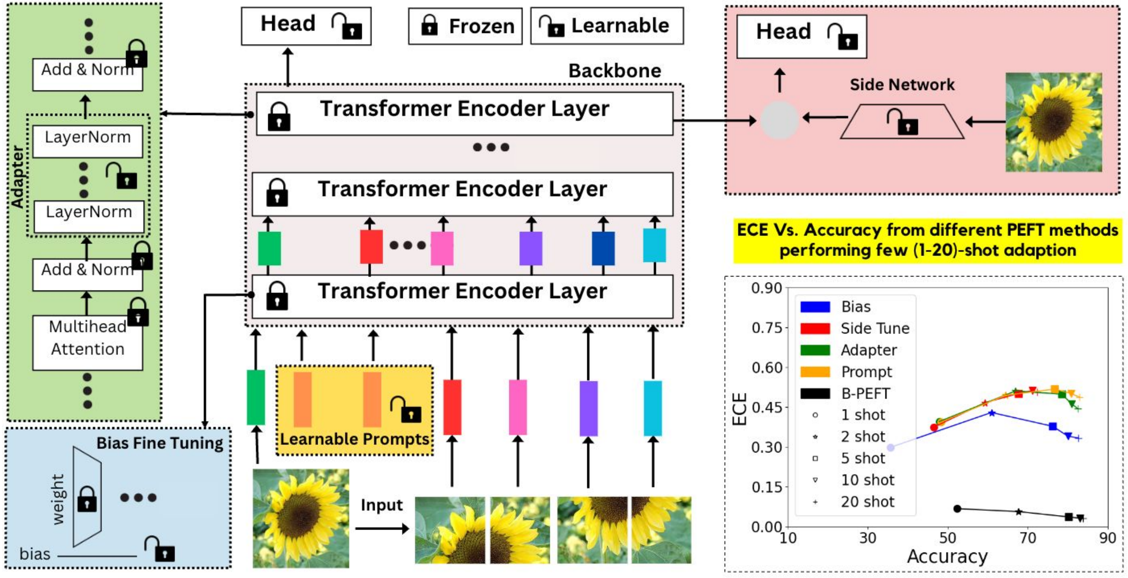
NEURAL INFORMATION
PROCESSING SYSTEMS

Be Confident in What You Know: Bayesian Parameter Efficient Fine-Tuning of Vision Foundation Models

*Deep Shankar Pandey†, Spandan Pyakurel†, Qi Yu**

Rochester Institute of Technology

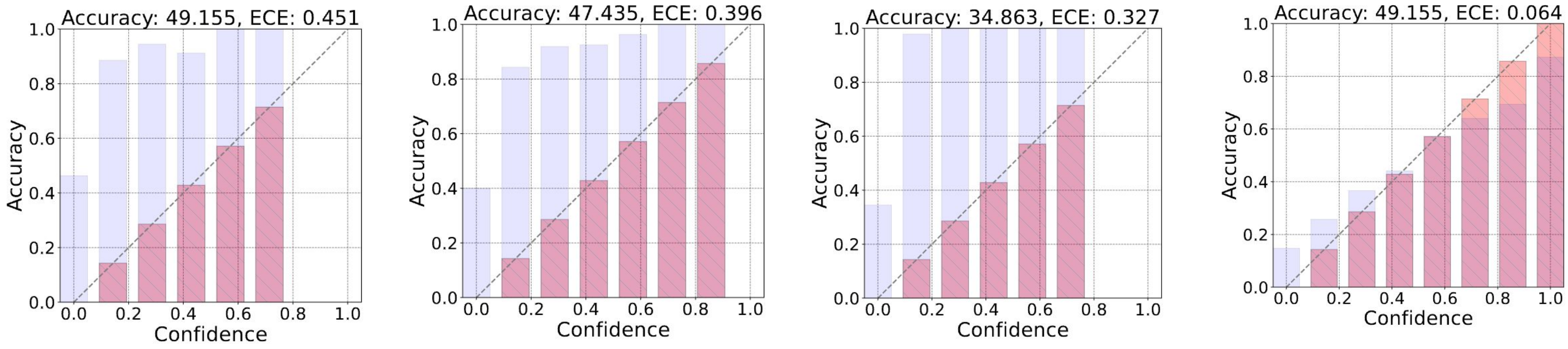
Few shot adaptation



(a) (b)

Accuracy Vs. ECE on CIFAR100 few-shot adaptation from different PEFT methods

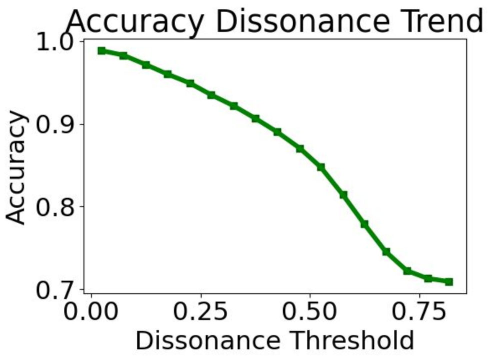
Under-Confidence Issue



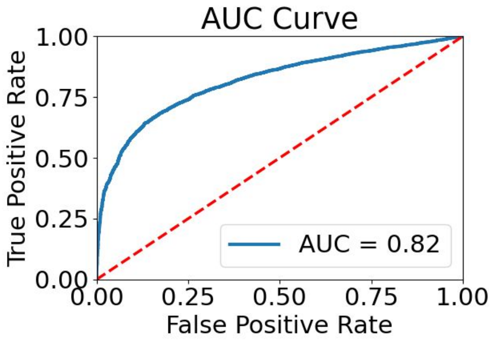
(a) Prompt *(b) Adapter* *(c) Bias* *(d) Proposed Method*

Why Is The Model Accurate?

- ❑ Low dissonance implies a high accuracy, shown by **Dissonance-Accuracy curve.**
- ❑ Area Under the Curve of the Accuracy vs. (1 − dissonance) is high.
- ❑ Model is able to clearly discriminate the ground-truth label from the rest without much confusion.



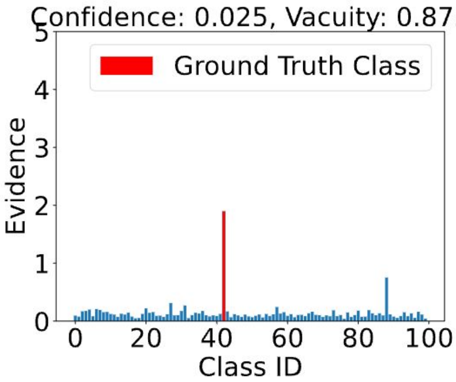
(a) Dissonance-Accuracy



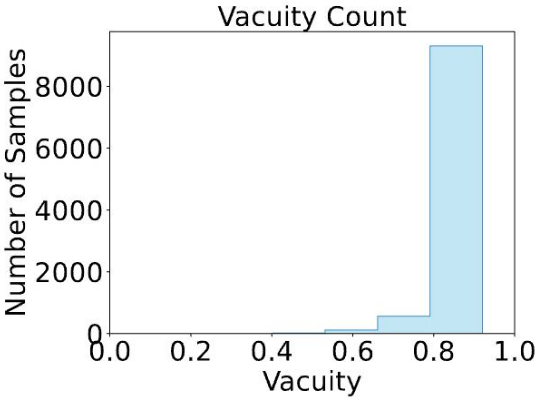
(b) AUC Curve

Why Is The Model Under-Confident?

- Model generally assigns very low evidence to all the labels, including the correct one.
- Higher vacuity assigned to the most of the samples due to lack of evidence.



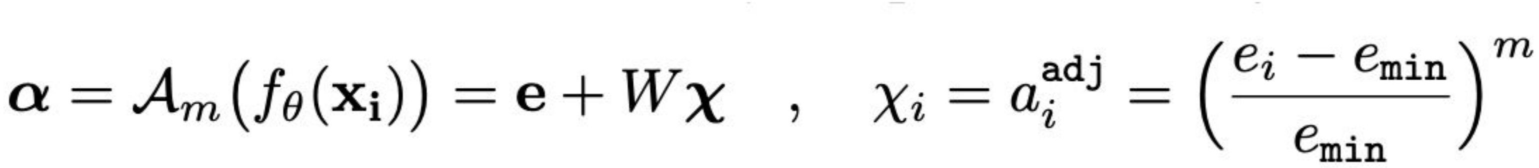
(a) Evidence of classes



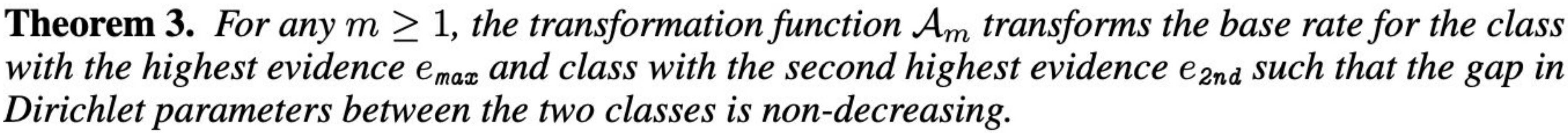
(b) Vacuity Distribution

Base Rate Adjustment

- Adjust the prior belief gained through pre-training
- The relative order of the Dirichlet parameters assigned to different classes is preserved
- The gap between the Dirichlet parameters for different classes is transformed such that the model becomes more confident in its predictions



$$\boldsymbol{\alpha} = \mathcal{A}_m(f_\theta(\mathbf{x_i})) = \mathbf{e} + W\boldsymbol{\chi} \quad , \quad \chi_i = a_i^{\mathtt{adj}} = \left(\frac{e_i - e_{\mathtt{min}}}{e_{\mathtt{min}}}\right)^m$$



Theorem 3. *For any $m \geq 1$, the transformation function $\mathcal{A}_m$ transforms the base rate for the class with the highest evidence e_{max} and class with the second highest evidence e_{2nd} such that the gap in Dirichlet parameters between the two classes is non-decreasing.*

Building A Diversity Induced Evidential Ensemble

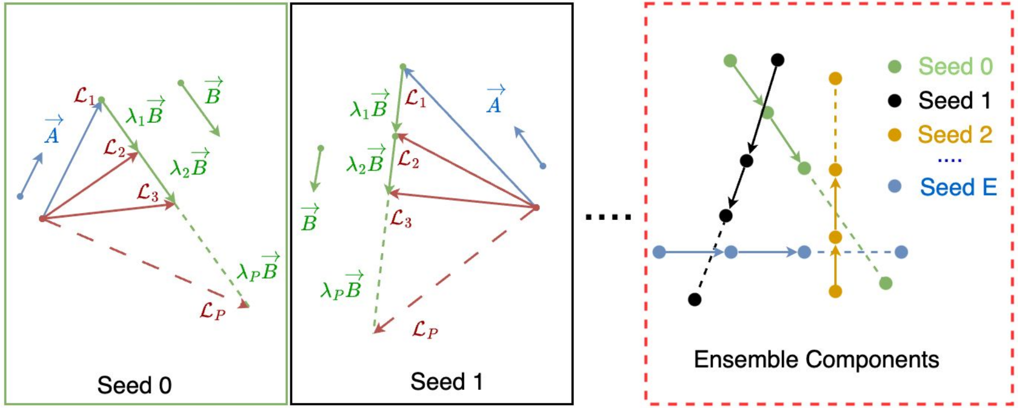
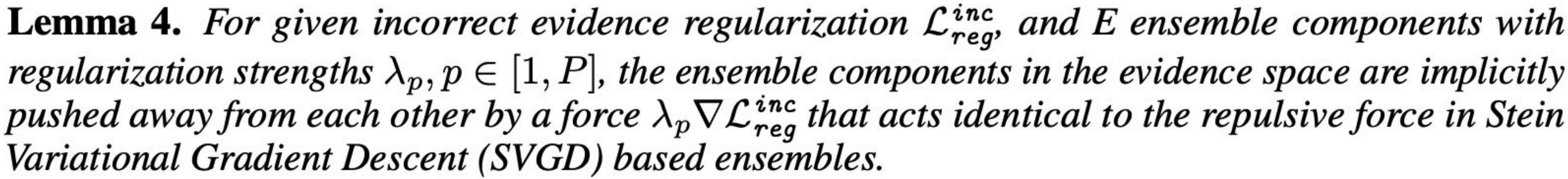
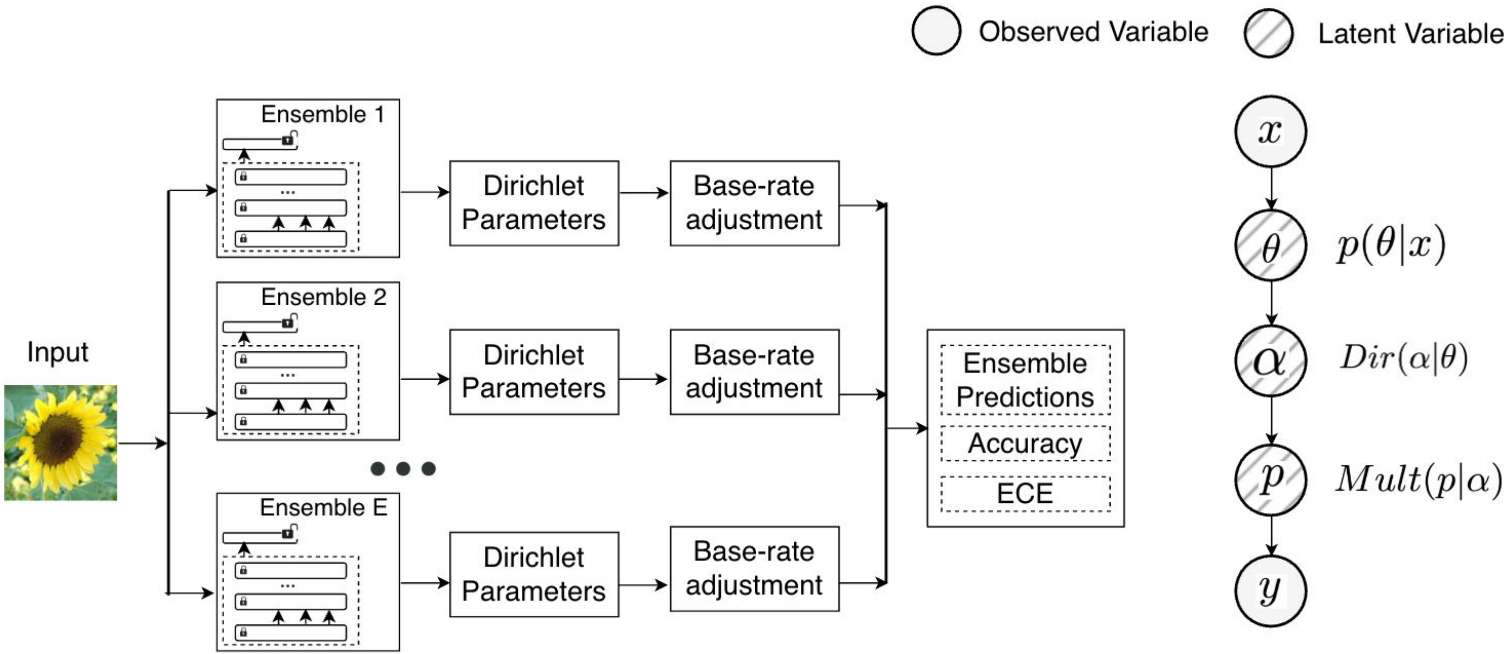


Illustration of diversity



Lemma 4. *For given incorrect evidence regularization $\mathcal{L}^{inc}_{reg}$, and E ensemble components with regularization strengths $\lambda_p, p \in [1, P]$, the ensemble components in the evidence space are implicitly pushed away from each other by a force $\lambda_p \nabla \mathcal{L}^{inc}_{reg}$ that acts identical to the repulsive force in Stein Variational Gradient Descent (SVGD) based ensembles.*

Overview



Conceptual diagram

Experimental Results

Standard PEFT Methods

- ☐ Reasonable generalization performance
- ☐ **Poor ECE (underconfidence issue)**

Base-rate adjustment

- ☐ Addresses underconfidence issue

B-PEFT

- ☐ **Further improves calibration (Best ECE)**
- ☐ **Superior generalization performance**

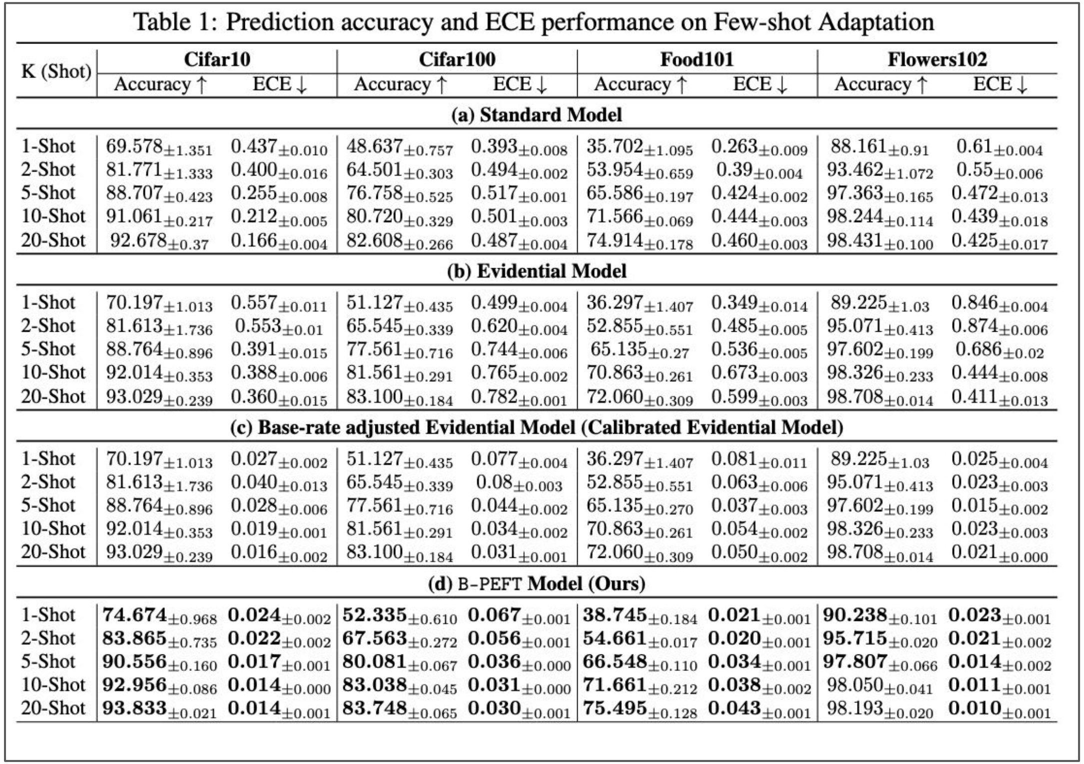


Table 1: Prediction accuracy and ECE performance on Few-shot Adaptation

K (Shot)	**Cifar10**		**Cifar100**		**Food101**		**Flowers102**	
	Accuracy ↑	ECE ↓	Accuracy ↑	ECE ↓	Accuracy ↑	ECE ↓	Accuracy ↑	ECE ↓
(a) Standard Model								
1-Shot	$69.578_{\pm1.351}$	$0.437_{\pm0.010}$	$48.637_{\pm0.757}$	$0.393_{\pm0.008}$	$35.702_{\pm1.095}$	$0.263_{\pm0.009}$	$88.161_{\pm0.91}$	$0.61_{\pm0.004}$
2-Shot	$81.771_{\pm1.333}$	$0.400_{\pm0.016}$	$64.501_{\pm0.303}$	$0.494_{\pm0.002}$	$53.954_{\pm0.659}$	$0.39_{\pm0.004}$	$93.462_{\pm1.072}$	$0.55_{\pm0.006}$
5-Shot	$88.707_{\pm0.423}$	$0.255_{\pm0.008}$	$76.758_{\pm0.525}$	$0.517_{\pm0.001}$	$65.586_{\pm0.197}$	$0.424_{\pm0.002}$	$97.363_{\pm0.165}$	$0.472_{\pm0.013}$
10-Shot	$91.061_{\pm0.217}$	$0.212_{\pm0.005}$	$80.720_{\pm0.329}$	$0.501_{\pm0.003}$	$71.566_{\pm0.069}$	$0.444_{\pm0.003}$	$98.244_{\pm0.114}$	$0.439_{\pm0.018}$
20-Shot	$92.678_{\pm0.37}$	$0.166_{\pm0.004}$	$82.608_{\pm0.266}$	$0.487_{\pm0.004}$	$74.914_{\pm0.178}$	$0.460_{\pm0.003}$	$98.431_{\pm0.100}$	$0.425_{\pm0.017}$
(b) Evidential Model								
1-Shot	$70.197_{\pm1.013}$	$0.557_{\pm0.011}$	$51.127_{\pm0.435}$	$0.499_{\pm0.004}$	$36.297_{\pm1.407}$	$0.349_{\pm0.014}$	$89.225_{\pm1.03}$	$0.846_{\pm0.004}$
2-Shot	$81.613_{\pm1.736}$	$0.553_{\pm0.01}$	$65.545_{\pm0.339}$	$0.620_{\pm0.004}$	$52.855_{\pm0.551}$	$0.485_{\pm0.005}$	$95.071_{\pm0.413}$	$0.874_{\pm0.006}$
5-Shot	$88.764_{\pm0.896}$	$0.391_{\pm0.015}$	$77.561_{\pm0.716}$	$0.744_{\pm0.006}$	$65.135_{\pm0.27}$	$0.536_{\pm0.005}$	$97.602_{\pm0.199}$	$0.686_{\pm0.02}$
10-Shot	$92.014_{\pm0.353}$	$0.388_{\pm0.006}$	$81.561_{\pm0.291}$	$0.765_{\pm0.002}$	$70.863_{\pm0.261}$	$0.673_{\pm0.003}$	$98.326_{\pm0.233}$	$0.444_{\pm0.008}$
20-Shot	$93.029_{\pm0.239}$	$0.360_{\pm0.015}$	$83.100_{\pm0.184}$	$0.782_{\pm0.001}$	$72.060_{\pm0.309}$	$0.599_{\pm0.003}$	$98.708_{\pm0.014}$	$0.411_{\pm0.013}$
(c) Base-rate adjusted Evidential Model (Calibrated Evidential Model)								
1-Shot	$70.197_{\pm1.013}$	$0.027_{\pm0.002}$	$51.127_{\pm0.435}$	$0.077_{\pm0.004}$	$36.297_{\pm1.407}$	$0.081_{\pm0.011}$	$89.225_{\pm1.03}$	$0.025_{\pm0.004}$
2-Shot	$81.613_{\pm1.736}$	$0.040_{\pm0.013}$	$65.545_{\pm0.339}$	$0.08_{\pm0.003}$	$52.855_{\pm0.551}$	$0.063_{\pm0.006}$	$95.071_{\pm0.413}$	$0.023_{\pm0.003}$
5-Shot	$88.764_{\pm0.896}$	$0.028_{\pm0.006}$	$77.561_{\pm0.716}$	$0.044_{\pm0.002}$	$65.135_{\pm0.270}$	$0.037_{\pm0.003}$	$97.602_{\pm0.199}$	$0.015_{\pm0.002}$
10-Shot	$92.014_{\pm0.353}$	$0.019_{\pm0.001}$	$81.561_{\pm0.291}$	$0.034_{\pm0.002}$	$70.863_{\pm0.261}$	$0.054_{\pm0.002}$	$98.326_{\pm0.233}$	$0.023_{\pm0.003}$
20-Shot	$93.029_{\pm0.239}$	$0.016_{\pm0.002}$	$83.100_{\pm0.184}$	$0.031_{\pm0.001}$	$72.060_{\pm0.309}$	$0.050_{\pm0.002}$	$98.708_{\pm0.014}$	$0.021_{\pm0.000}$
(d) `B-PEFT` **Model (Ours)**								
1-Shot	$\mathbf{74.674}_{\pm0.968}$	$\mathbf{0.024}_{\pm0.002}$	$\mathbf{52.335}_{\pm0.610}$	$\mathbf{0.067}_{\pm0.001}$	$\mathbf{38.745}_{\pm0.184}$	$\mathbf{0.021}_{\pm0.001}$	$\mathbf{90.238}_{\pm0.101}$	$\mathbf{0.023}_{\pm0.001}$
2-Shot	$\mathbf{83.865}_{\pm0.735}$	$\mathbf{0.022}_{\pm0.002}$	$\mathbf{67.563}_{\pm0.272}$	$\mathbf{0.056}_{\pm0.001}$	$\mathbf{54.661}_{\pm0.017}$	$\mathbf{0.020}_{\pm0.001}$	$\mathbf{95.715}_{\pm0.020}$	$\mathbf{0.021}_{\pm0.002}$
5-Shot	$\mathbf{90.556}_{\pm0.160}$	$\mathbf{0.017}_{\pm0.001}$	$\mathbf{80.081}_{\pm0.067}$	$\mathbf{0.036}_{\pm0.000}$	$\mathbf{66.548}_{\pm0.110}$	$\mathbf{0.034}_{\pm0.001}$	$\mathbf{97.807}_{\pm0.066}$	$\mathbf{0.014}_{\pm0.002}$
10-Shot	$\mathbf{92.956}_{\pm0.086}$	$\mathbf{0.014}_{\pm0.000}$	$\mathbf{83.038}_{\pm0.045}$	$\mathbf{0.031}_{\pm0.000}$	$\mathbf{71.661}_{\pm0.212}$	$\mathbf{0.038}_{\pm0.002}$	$98.050_{\pm0.041}$	$\mathbf{0.011}_{\pm0.001}$
20-Shot	$\mathbf{93.833}_{\pm0.021}$	$\mathbf{0.014}_{\pm0.001}$	$\mathbf{83.748}_{\pm0.065}$	$\mathbf{0.030}_{\pm0.001}$	$\mathbf{75.495}_{\pm0.128}$	$\mathbf{0.043}_{\pm0.001}$	$98.193_{\pm0.020}$	$\mathbf{0.010}_{\pm0.001}$

Experimental Results

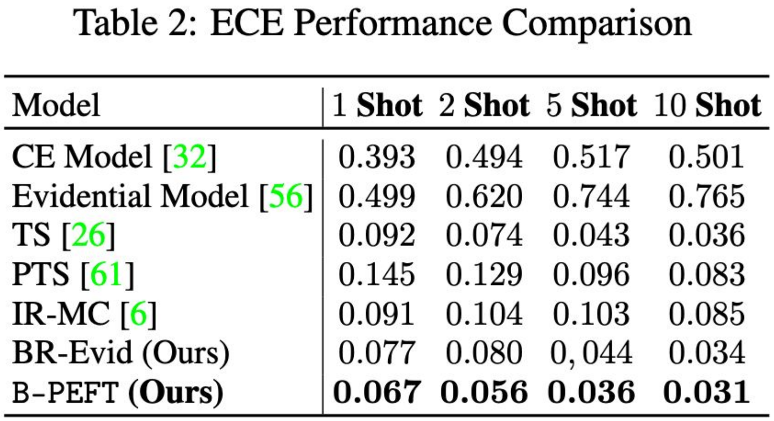


Table 2: ECE Performance Comparison

Model	1 **Shot**	2 **Shot**	5 **Shot**	10 **Shot**
CE Model [32]	0.393	0.494	0.517	0.501
Evidential Model [56]	0.499	0.620	0.744	0.765
TS [26]	0.092	0.074	0.043	0.036
PTS [61]	0.145	0.129	0.096	0.083
IR-MC [6]	0.091	0.104	0.103	0.085
BR-Evid (Ours)	0.077	0.080	0, 044	0.034
`B-PEFT` **(Ours)**	**0.067**	**0.056**	**0.036**	**0.031**

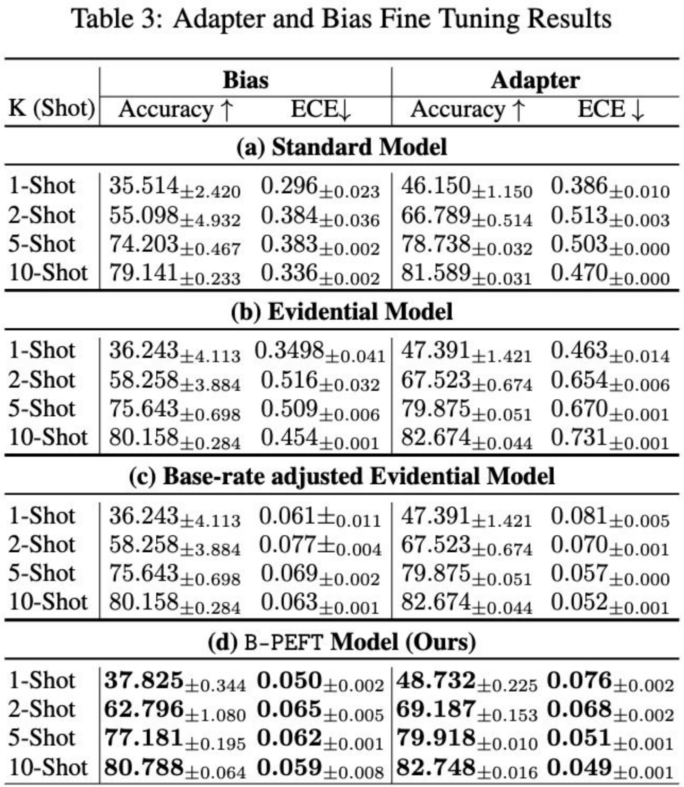
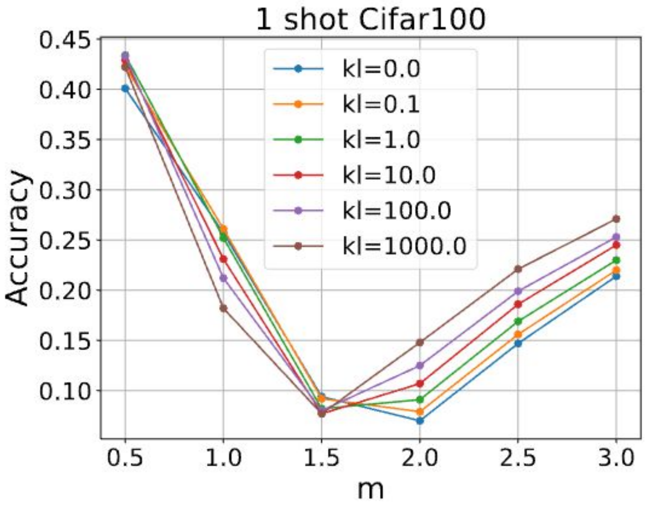


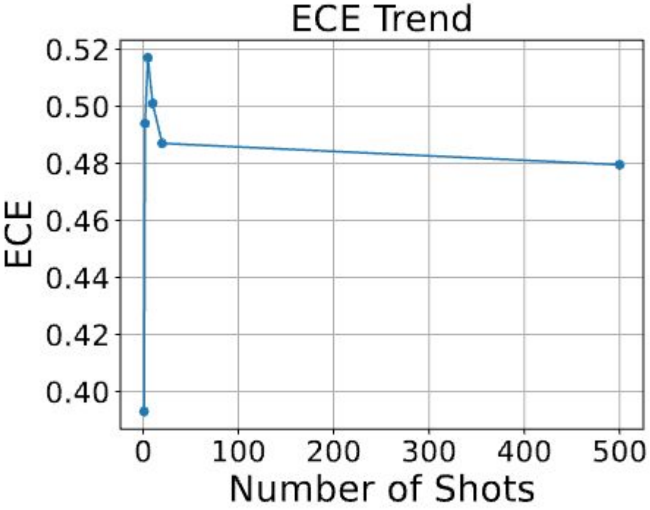
Table 3: Adapter and Bias Fine Tuning Results

	Bias		**Adapter**	
K (Shot)	Accuracy ↑	ECE↓	Accuracy ↑	ECE ↓
(a) Standard Model				
1-Shot	$35.514_{\pm 2.420}$	$0.296_{\pm 0.023}$	$46.150_{\pm 1.150}$	$0.386_{\pm 0.010}$
2-Shot	$55.098_{\pm 4.932}$	$0.384_{\pm 0.036}$	$66.789_{\pm 0.514}$	$0.513_{\pm 0.003}$
5-Shot	$74.203_{\pm 0.467}$	$0.383_{\pm 0.002}$	$78.738_{\pm 0.032}$	$0.503_{\pm 0.000}$
10-Shot	$79.141_{\pm 0.233}$	$0.336_{\pm 0.002}$	$81.589_{\pm 0.031}$	$0.470_{\pm 0.000}$
(b) Evidential Model				
1-Shot	$36.243_{\pm 4.113}$	$0.3498_{\pm 0.041}$	$47.391_{\pm 1.421}$	$0.463_{\pm 0.014}$
2-Shot	$58.258_{\pm 3.884}$	$0.516_{\pm 0.032}$	$67.523_{\pm 0.674}$	$0.654_{\pm 0.006}$
5-Shot	$75.643_{\pm 0.698}$	$0.509_{\pm 0.006}$	$79.875_{\pm 0.051}$	$0.670_{\pm 0.001}$
10-Shot	$80.158_{\pm 0.284}$	$0.454_{\pm 0.001}$	$82.674_{\pm 0.044}$	$0.731_{\pm 0.001}$
(c) Base-rate adjusted Evidential Model				
1-Shot	$36.243_{\pm 4.113}$	$0.061 \pm_{0.011}$	$47.391_{\pm 1.421}$	$0.081_{\pm 0.005}$
2-Shot	$58.258_{\pm 3.884}$	$0.077 \pm_{0.004}$	$67.523_{\pm 0.674}$	$0.070_{\pm 0.001}$
5-Shot	$75.643_{\pm 0.698}$	$0.069_{\pm 0.002}$	$79.875_{\pm 0.051}$	$0.057_{\pm 0.000}$
10-Shot	$80.158_{\pm 0.284}$	$0.063_{\pm 0.001}$	$82.674_{\pm 0.044}$	$0.052_{\pm 0.001}$
(d) `B-PEFT` **Model (Ours)**				
1-Shot	$\mathbf{37.825}_{\pm 0.344}$	$\mathbf{0.050}_{\pm 0.002}$	$\mathbf{48.732}_{\pm 0.225}$	$\mathbf{0.076}_{\pm 0.002}$
2-Shot	$\mathbf{62.796}_{\pm 1.080}$	$\mathbf{0.065}_{\pm 0.005}$	$\mathbf{69.187}_{\pm 0.153}$	$\mathbf{0.068}_{\pm 0.002}$
5-Shot	$\mathbf{77.181}_{\pm 0.195}$	$\mathbf{0.062}_{\pm 0.001}$	$\mathbf{79.918}_{\pm 0.010}$	$\mathbf{0.051}_{\pm 0.001}$
10-Shot	$\mathbf{80.788}_{\pm 0.064}$	$\mathbf{0.059}_{\pm 0.008}$	$\mathbf{82.748}_{\pm 0.016}$	$\mathbf{0.049}_{\pm 0.001}$

Experimental Results



Impact of hyperparameter



Number of Shots vs ECE

Summary

- ❏ Standard PEFT techniques for supervised vision foundation transformer models lead to accurate but **poorly calibrated and highly underconfident models**
- ❏ **B-PEFT:** Enables uncertainty awareness with improved generalization and calibration.

Thank you!

All Codes *available at* https://github.com/ritmininglab/B-PEFT

Any Questions? *Feel free to reach out to us at Mining Lab RIT (https://www.rit.edu/mining/)*