GIST AILAB

NEURAL INFORMATION PROCESSING SYSTEMS

Curriculum Fine-tuning of Vision Foundation Model for Medical Image Classification Under Label Noise

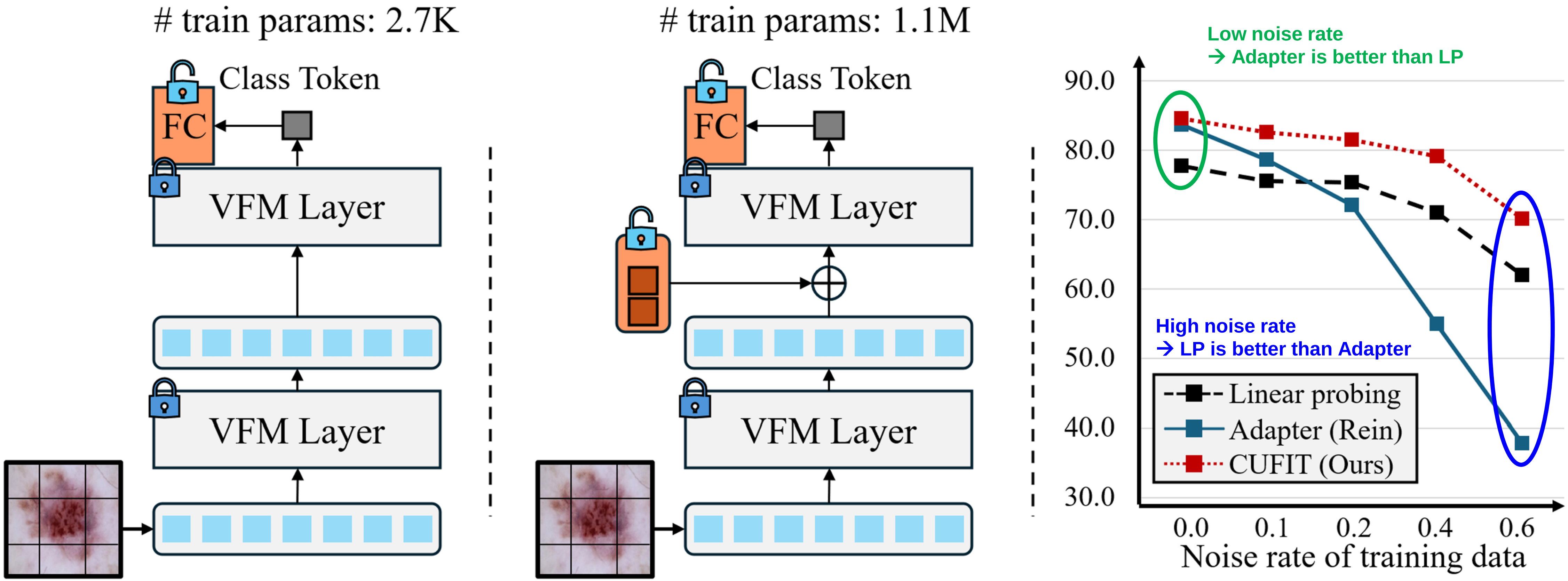
Yeonguk Yu · Minhwan Ko · Sungho Shin · Kangmin Kim · Kyoobin Lee

Artificial Intelligence Lab

School of Integrated Technology

Gwangju Institute of Science and Technology (GIST)

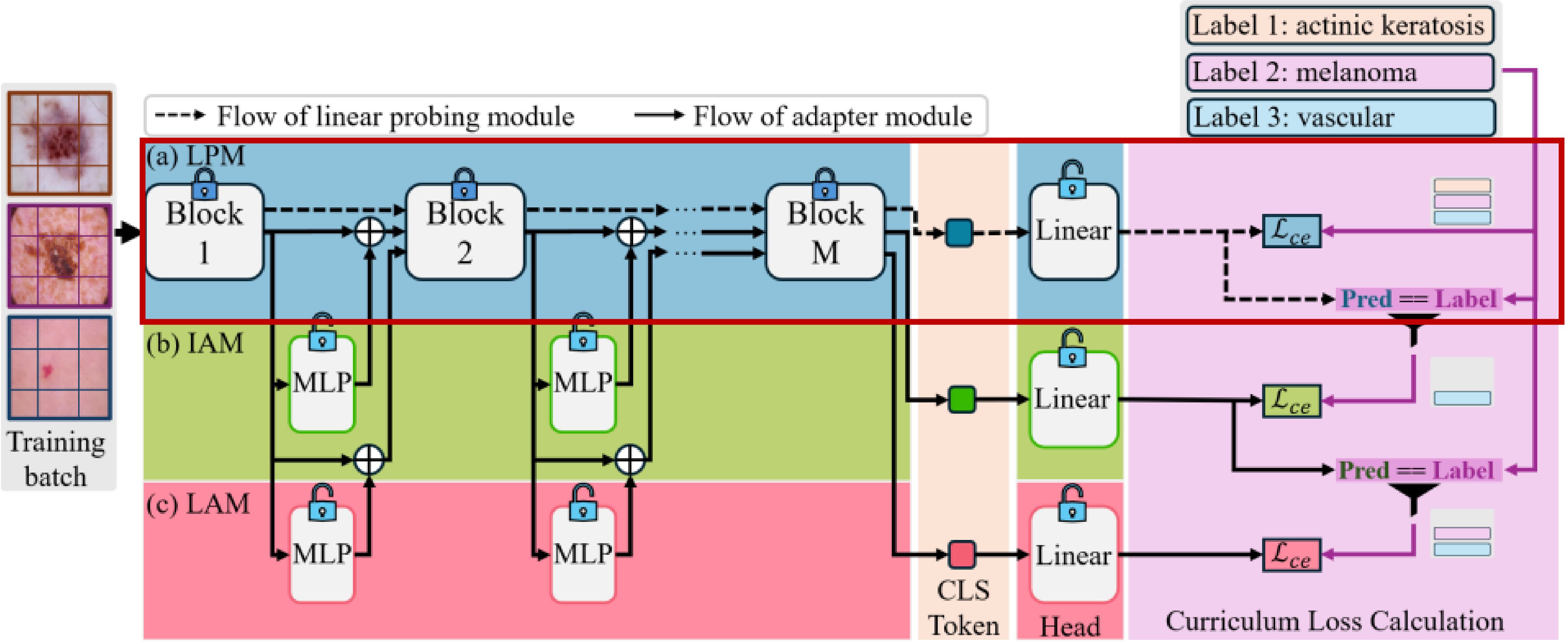
Introduction: Learning with Noisy Label with Vision Foundation Model



(a) Linear probing (b) Using adapters (c) Performance comparison

- The fine-tuning paradigms of vision foundation models (a and b) have both advantages and disadvantages when trained on noisy data
- The proposed fine-tuning paradigm, CUFIT, consists of one linear probing and two adapters to combat noisy labels during the training stage.

Method: Curriculum Fine-tuning (CUFIT)



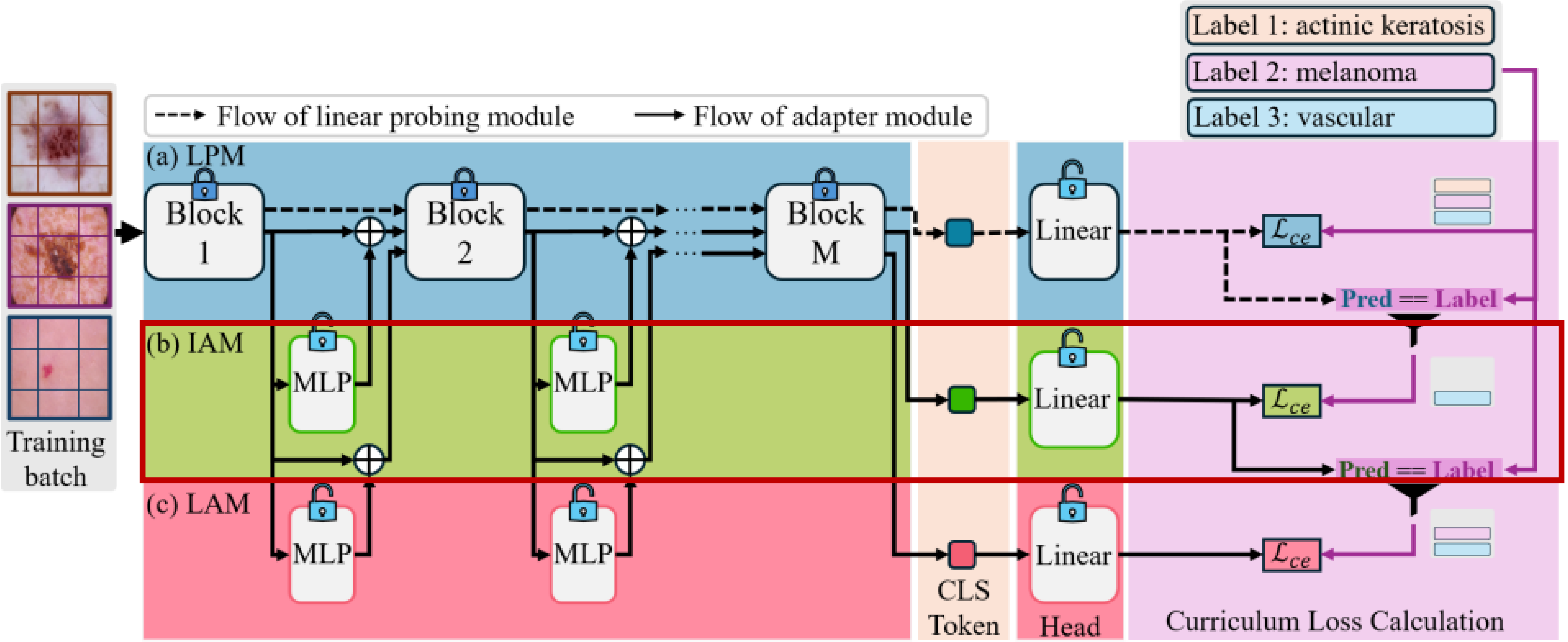
1. The Linear Probing Module (LPM) is robust against noisy labels

→ All training samples are used to train LPM

→ We select training samples when the LPM's predictions are equal to the annotations (First Filtering)

→ The filtered dataset has <u>**a low noise rate but contains only a few samples**</u>

Method: Curriculum Fine-tuning (CUFIT)



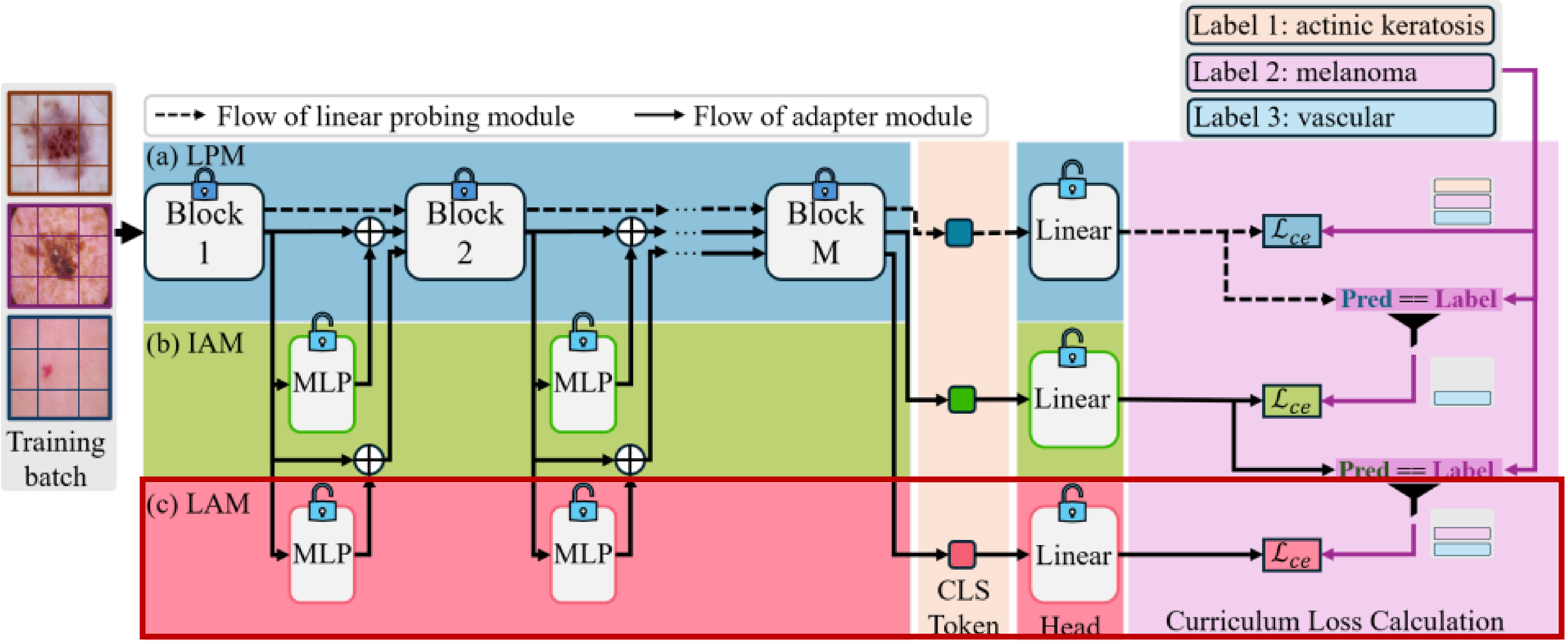
2. The adapter module outperforms linear probing when the noise rate is low

→ The filtered dataset is used to train Intermediate Adapter Module (IAM)

→ We select samples where the IAM's predictions match the ground truth from the entire dataset (Second Filtering)

→ The second-filtered dataset has **a low noise rate and contains more samples**.

Method: **Cu**rriculum **Fi**ne-**t**uning (**CUFIT**)



3. The adapter module outperforms linear probing when the noise rate is low (better with more samples)

→ The second-filtered dataset is used to train the Last Adapter Module (LAM)

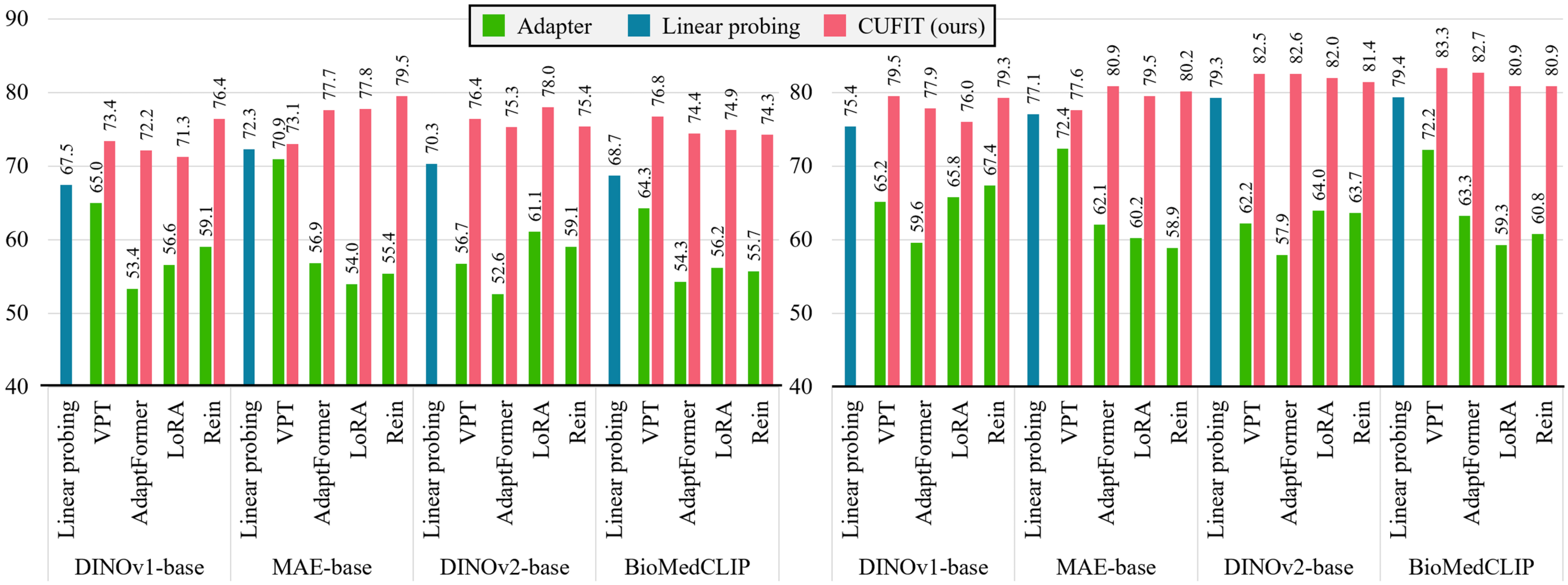
→ <u>**Only LAM is used during the inference stage**</u>

Experiment

	Dataset	Noise rate	Method		Previous methods of LNL			
			Linear probing	Rein	Co-teaching	JoCor	CoDis	Cufit (ours)
Skin disease classification	HAM10000	0.1	75.6	78.6	81.5	81.1	81.9	**82.6**
		0.2	75.3	72.1	79.1	79.4	80.1	**81.5**
		0.4	71.0	54.9	74.3	73.9	74.1	**79.1**
		0.6	61.9	37.8	67.3	67.1	66.1	**70.1**
		Mean	71.0	60.8	75.5	75.4	75.5	**78.3**
Diabetic retinopathy classification	APTOS-2019	0.1	79.2	82.5	82.8	**84.8**	83.2	84.2
		0.2	79.4	78.7	81.2	83.1	82.0	**84.2**
		0.4	79.5	77.2	79.5	76.0	79.5	**81.6**
		0.6	66.9	42.0	72.9	74.2	75.7	**76.3**
		Mean	76.3	68.9	79.1	79.5	80.1	**81.6**
Cell type classification	BloodMnist	0.1	97.2	95.9	98.6	98.5	98.5	**99.0**
		0.2	96.7	89.0	97.6	97.3	97.2	**98.8**
		0.4	95.8	69.3	93.7	93.0	93.5	**98.3**
		0.6	90.3	45.6	88.7	87.3	88.0	**98.2**
		Mean	95.0	75.0	94.7	94.0	94.3	**98.6**
Organ type classification	OrgancMnist	0.1	83.3	87.4	92.1	92.1	92.1	**93.7**
		0.2	82.9	82.0	90.9	91.9	90.7	**93.6**
		0.4	79.9	63.8	85.8	85.3	85.8	**91.6**
		0.6	72.2	43.1	82.8	82.6	81.9	**87.4**
		Mean	79.6	69.1	87.9	88.0	87.6	**91.6**

Table 1: Average test accuracy (%) on four simulated noisy datasets with different noise levels. The test accuracy is calculated over the last ten epochs. The best result and second best result in each case are highlighted in **bold** and underline, respectively

Discussion



(a) HAM10000 with 40% noise rate (b) APTOS-2019 with 40% noise rate

Conclusion

- Based on the insight that linear probing is robust to noisy label samples and adapter can generalize well with a few clean samples, we propose CUFIT, a Curriculum Fine-tuning paradigm for Vision Foundation Model for multi-class medical image classification.

- CUFIT utilizes three modules: a linear probing module and two adapters, to combat noisy labels during training stage. Specifically, CUFIT trains each module with filtered data in following order: linear probing → intermediate adapter → last adapter.

- Experiments demonstrate that CUFIT significantly improves the performance of VFMs in the presence of noisy labels for medical image classification. Also, we validate that CUFIT works well with various architectures showing that the method is model-agnostic.

Project page: