Entropy Minimization without Model Collapse: Mitigating Prediction Bias in Medical Imaging

Tim Nielen*[1], Sameer Ambekar*[1,2,4,5], Daniel M. Lang#[1,2], Julia A. Schnabel#[1,2,3,4,5]

[1]TU Munich [2]Helmholtz Munich [3]King's College London

[4]Munich Center for Machine Learning (MCML) [5]relAI - Konrad Zuse School of Excellence in Reliable AI

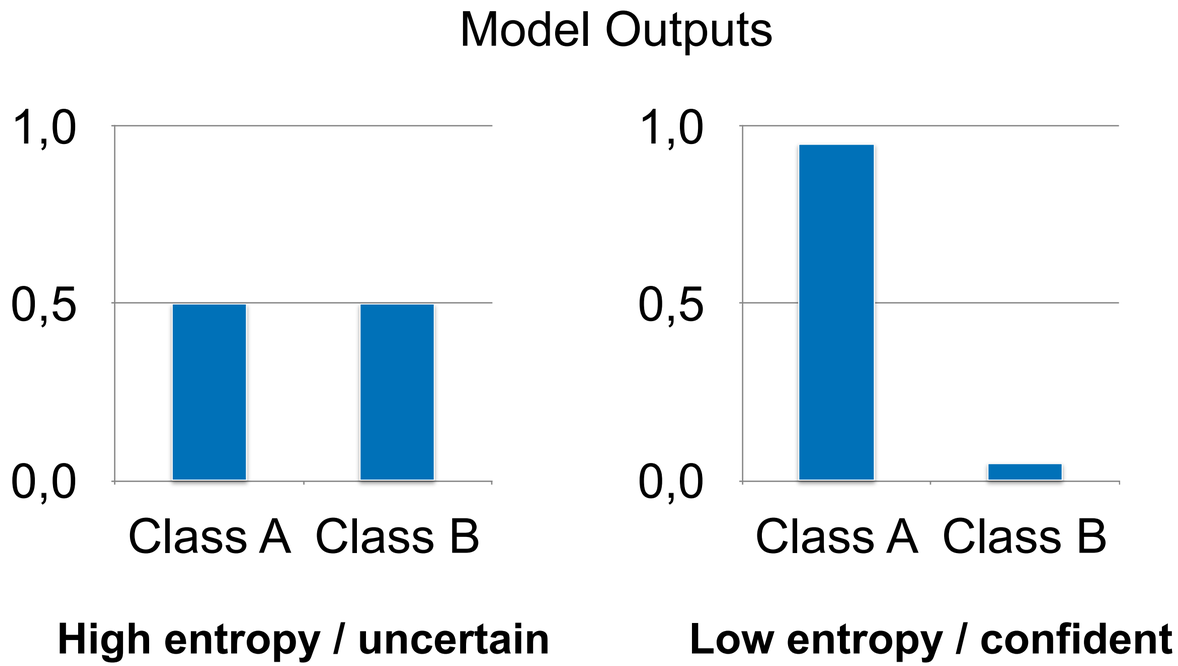
***Shared first-authorship #Shared last-authorship**

✨**NeurIPS 2026 Spotlight paper, Main track**✨

With funding from the:

Entropy Minimization (EM) and Sharpening

$$\theta^* = \arg\min_{\theta} \left(-\sum_{c=1}^{C} \hat{y}_c \log \hat{y}_c \right)$$



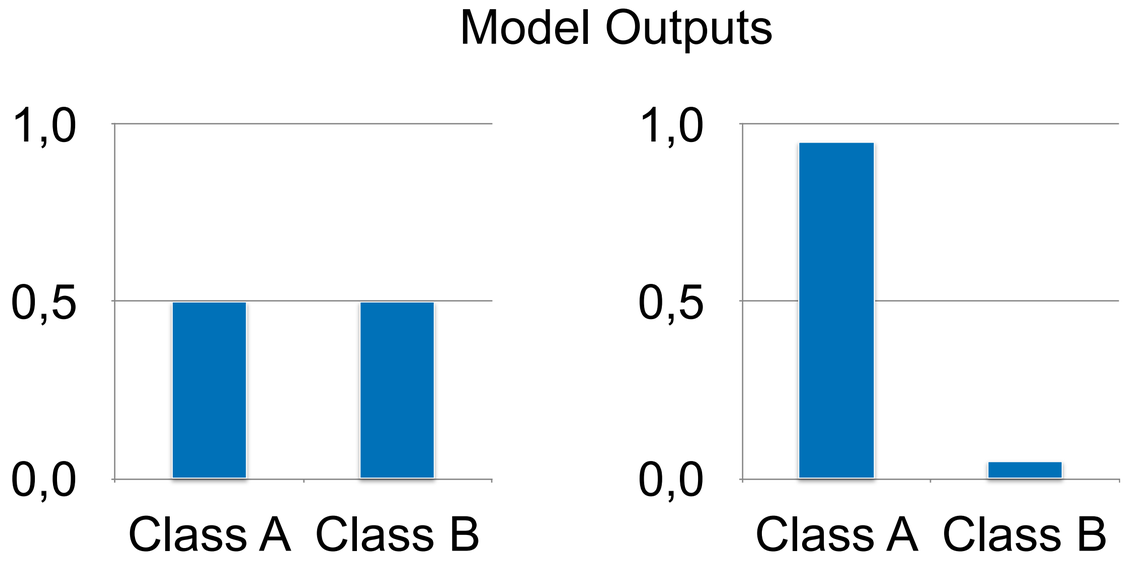
High entropy / uncertain

Low entropy / confident

Wang, Dequan, et al. "Tent: Fully Test-Time Adaptation by Entropy Minimization." *ICLR*, 2021.

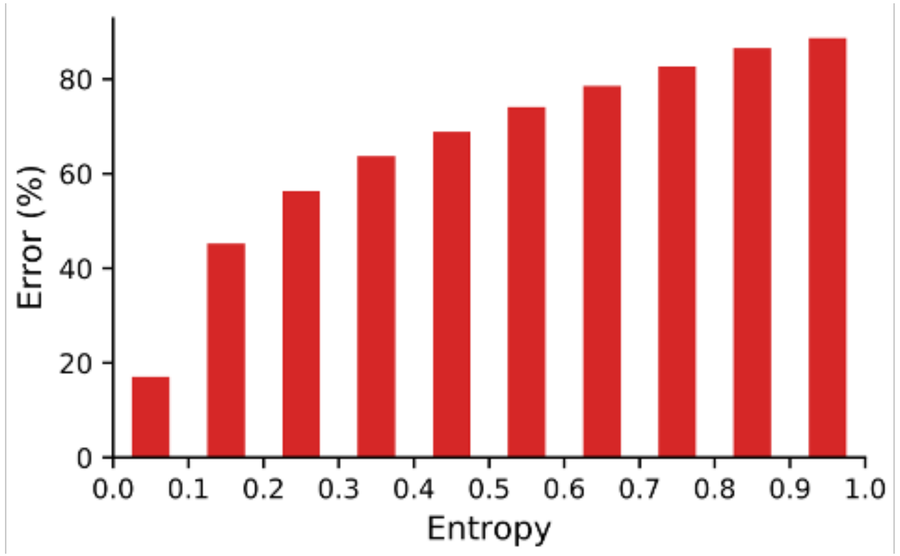
Entropy Minimization (EM) and Sharpening

$$\theta^* = \arg\min_{\theta} \left(-\sum_{c=1}^{C} \hat{y}_c \log \hat{y}_c \right)$$



High entropy / uncertain

Low entropy / confident

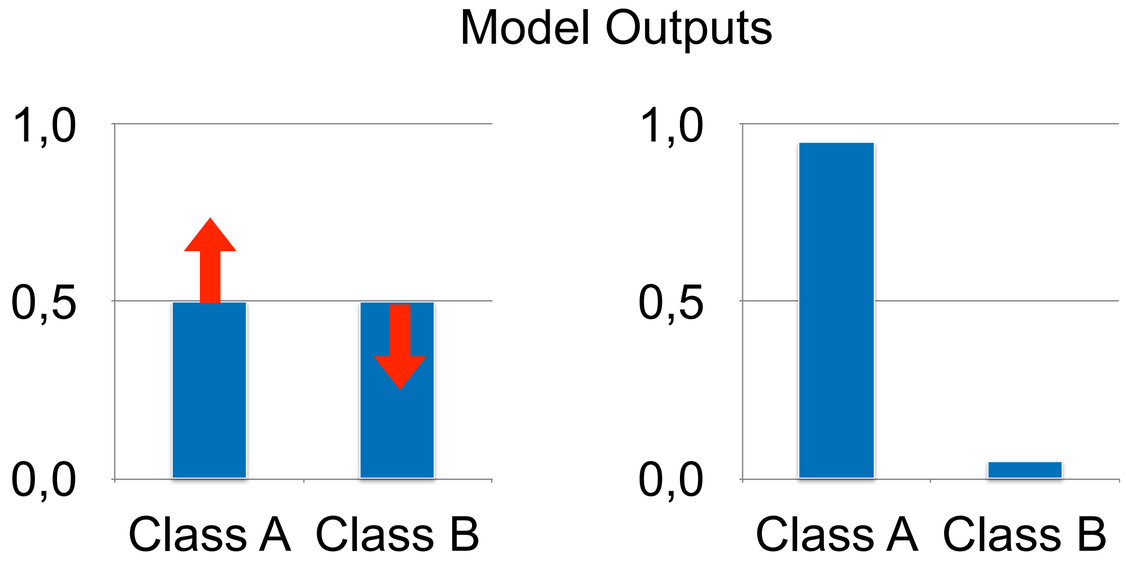


Predictions with low entropy have lower error rate

Wang, Dequan, et al. "Tent: Fully Test-Time Adaptation by Entropy Minimization." *ICLR*, 2021.

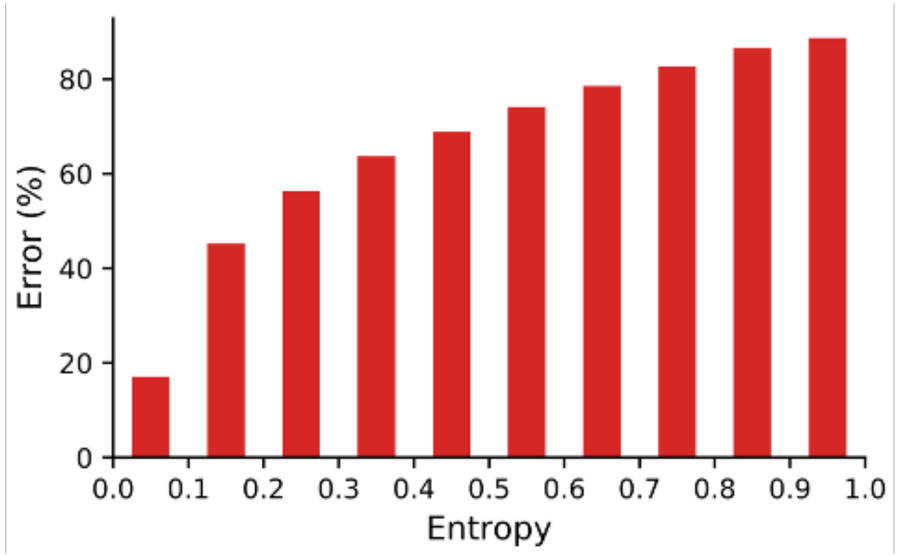
Entropy Minimization (EM) and Sharpening

$$\theta^* = \arg\min_{\theta} \left(-\sum_{c=1}^{C} \hat{y}_c \log \hat{y}_c \right)$$



High entropy / uncertain

Low entropy / confident

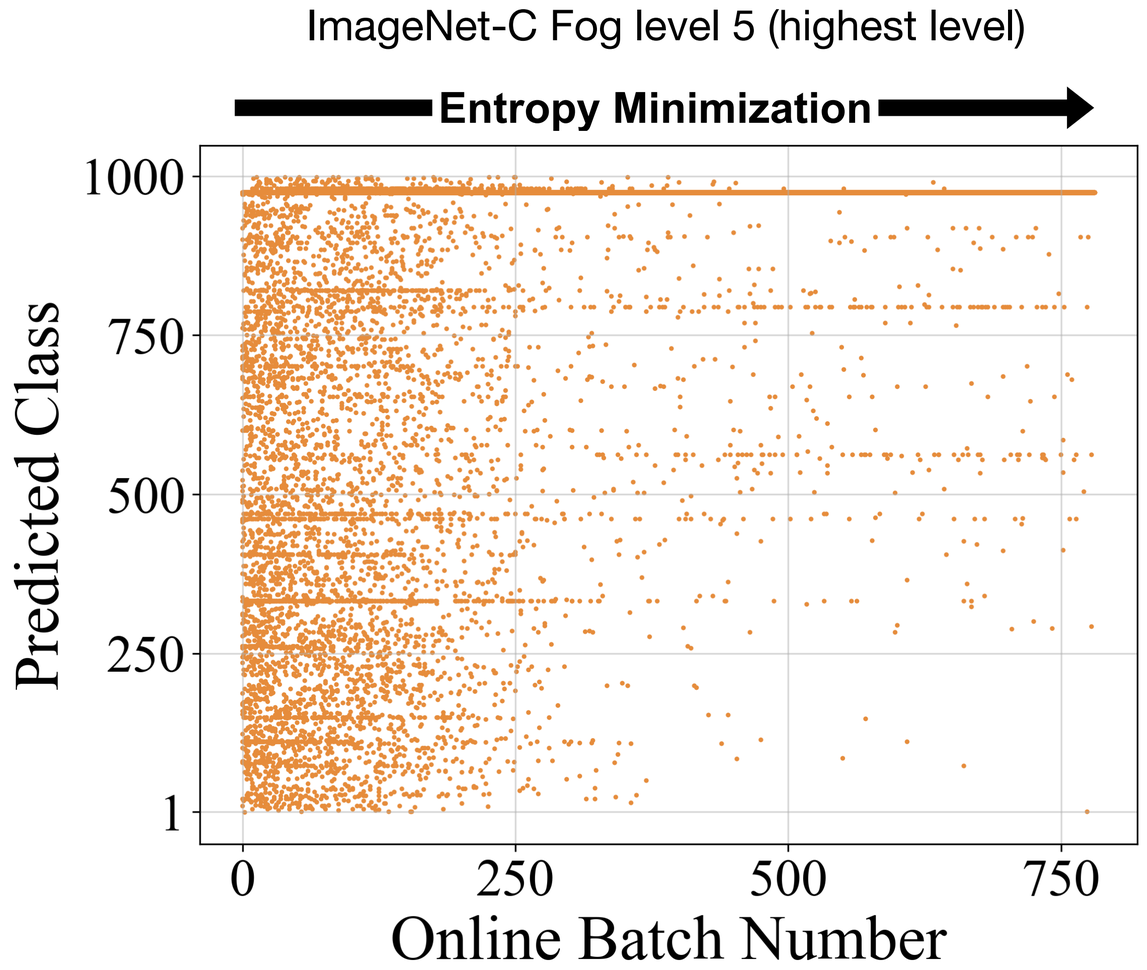


Predictions with low entropy have lower error rate

→ Use entropy as an unsupervised loss for domain adaptation at test-time

Wang, Dequan, et al. "Tent: Fully Test-Time Adaptation by Entropy Minimization." *ICLR*, 2021.

Entropy Minimization Causes Model Collapse

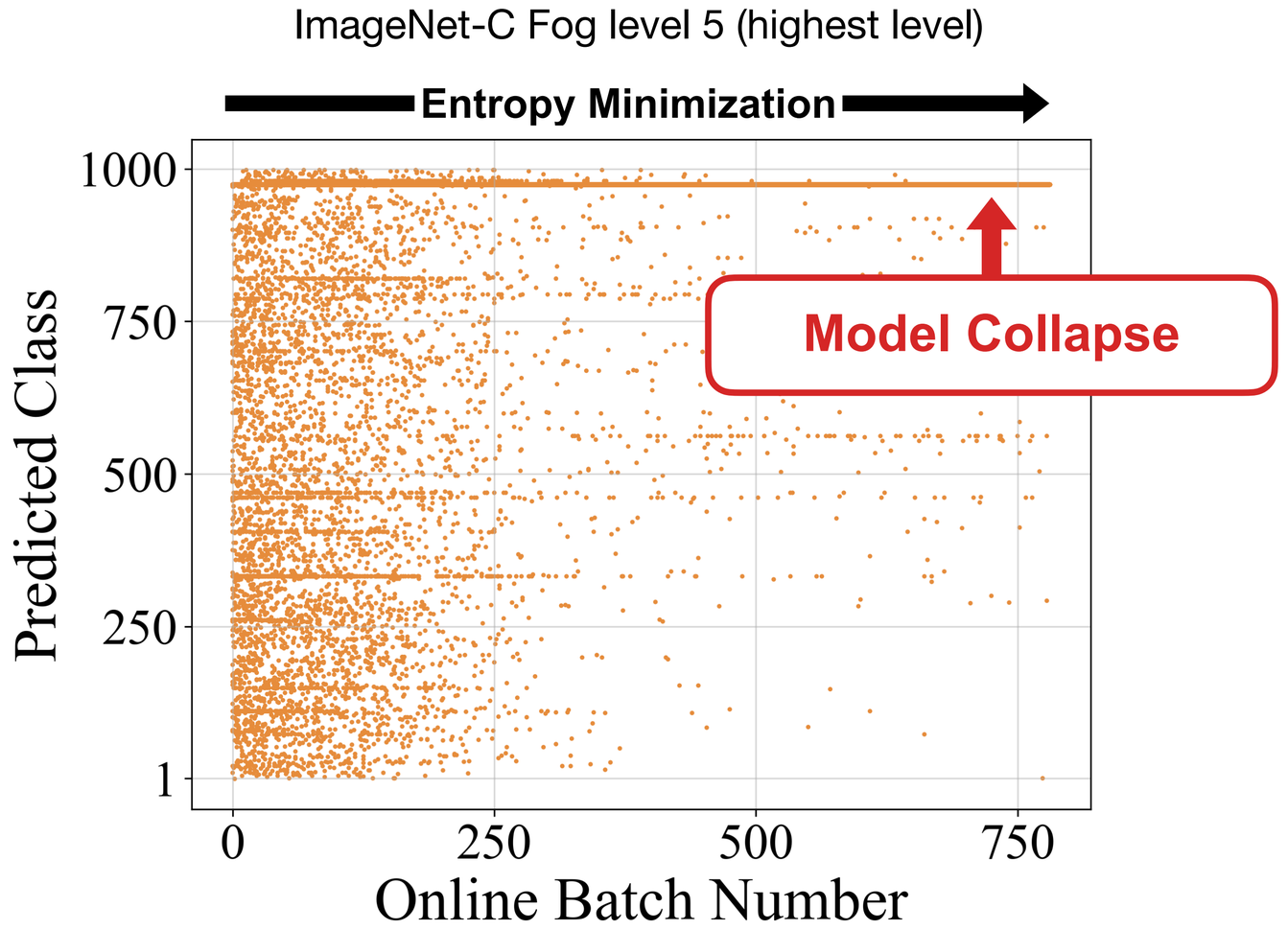


Model collapse: the model predicts the same class for all inputs

Niu, Shuaicheng, et al. "Towards Stable Test-time Adaptation in Dynamic Wild World." *ICLR*, 2023.

Entropy Minimization Causes Model Collapse

Indirect fixes (e.g., teacher-student models) are not enough



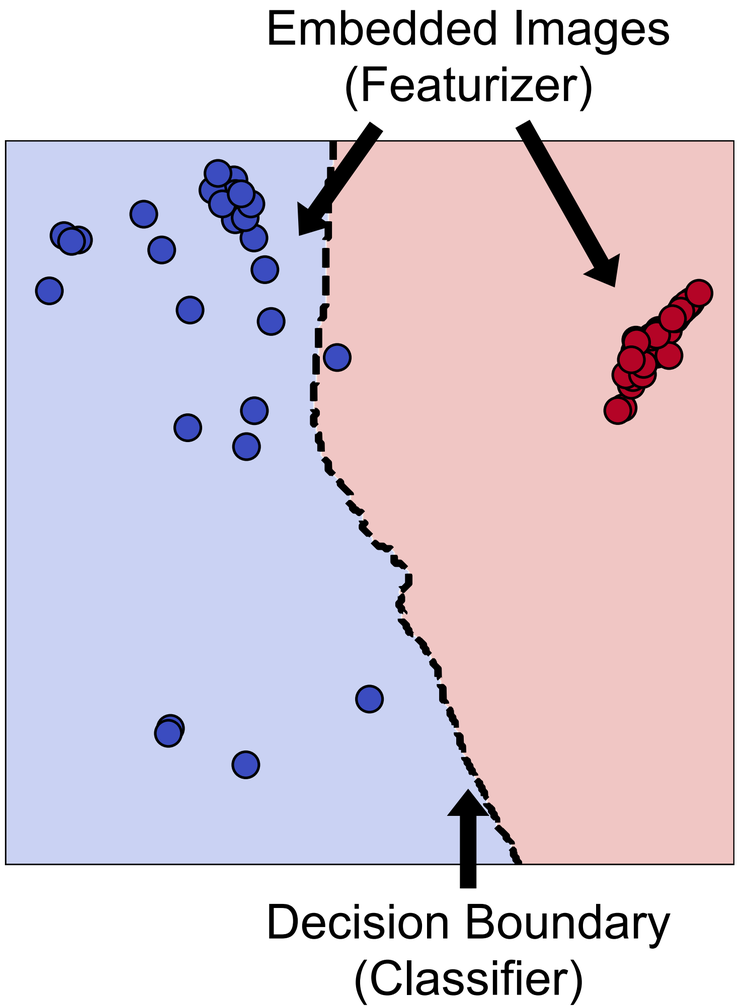
Collapse can be hard to detect without labels

Model collapse: the model predicts the same class for all inputs

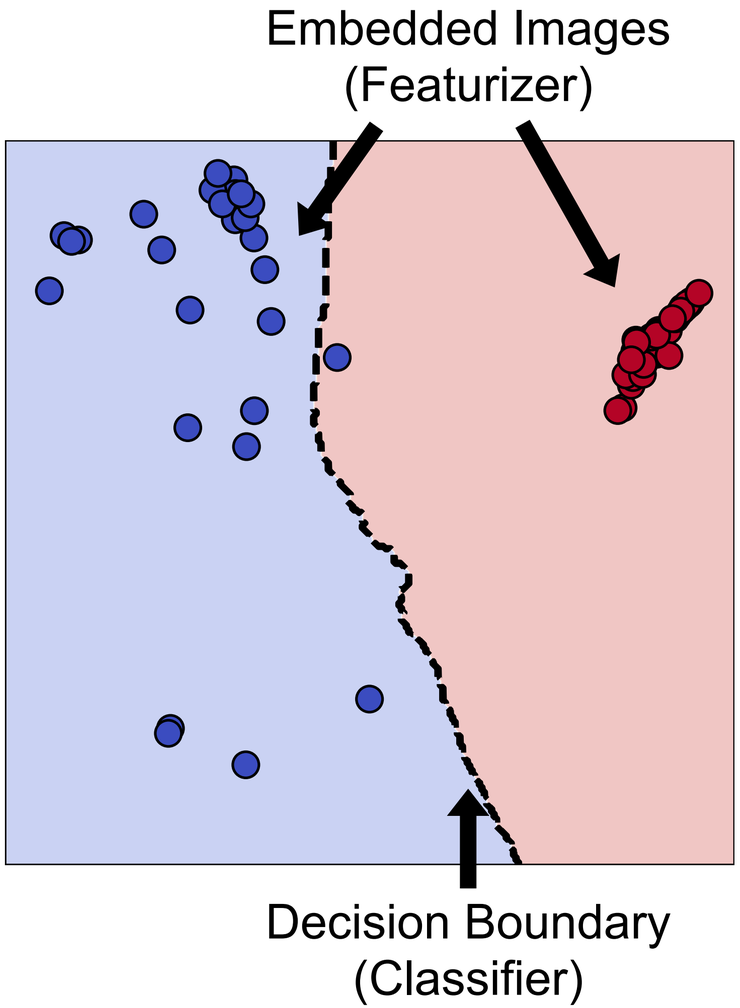
Niu, Shuaicheng, et al. "Towards Stable Test-time Adaptation in Dynamic Wild World." *ICLR*, 2023.

Why does Entropy Minimization lead to Model Collapse?

Feature Space Analysis



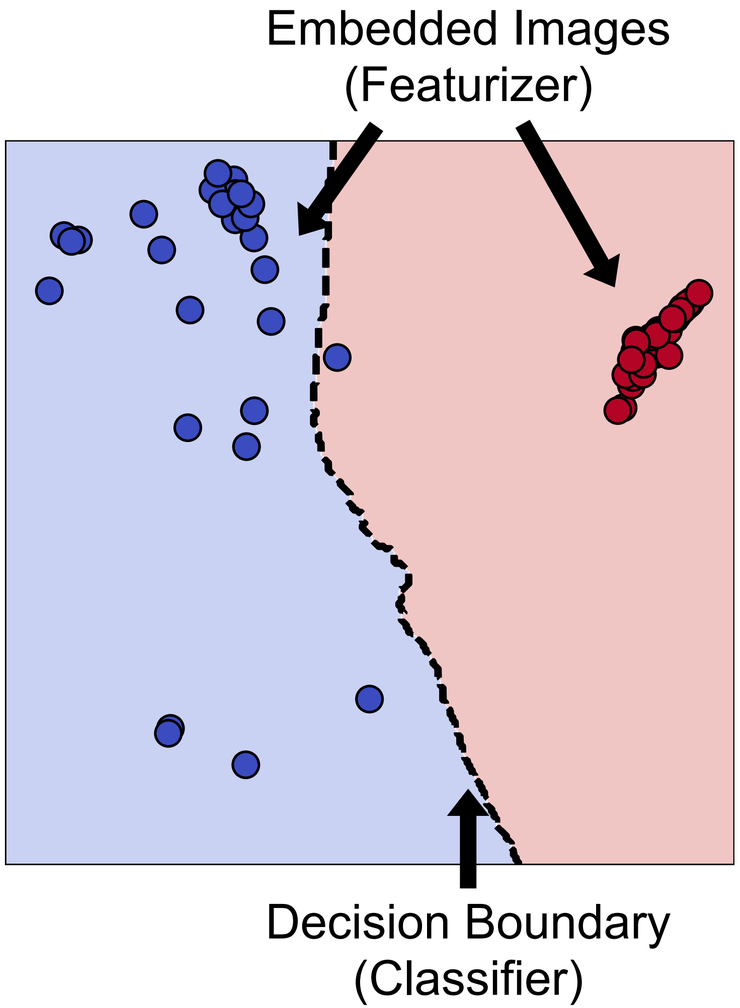
Feature Space Analysis



The classifier acts as a decision boundary in feature space

→ **stays fixed** during domain shift

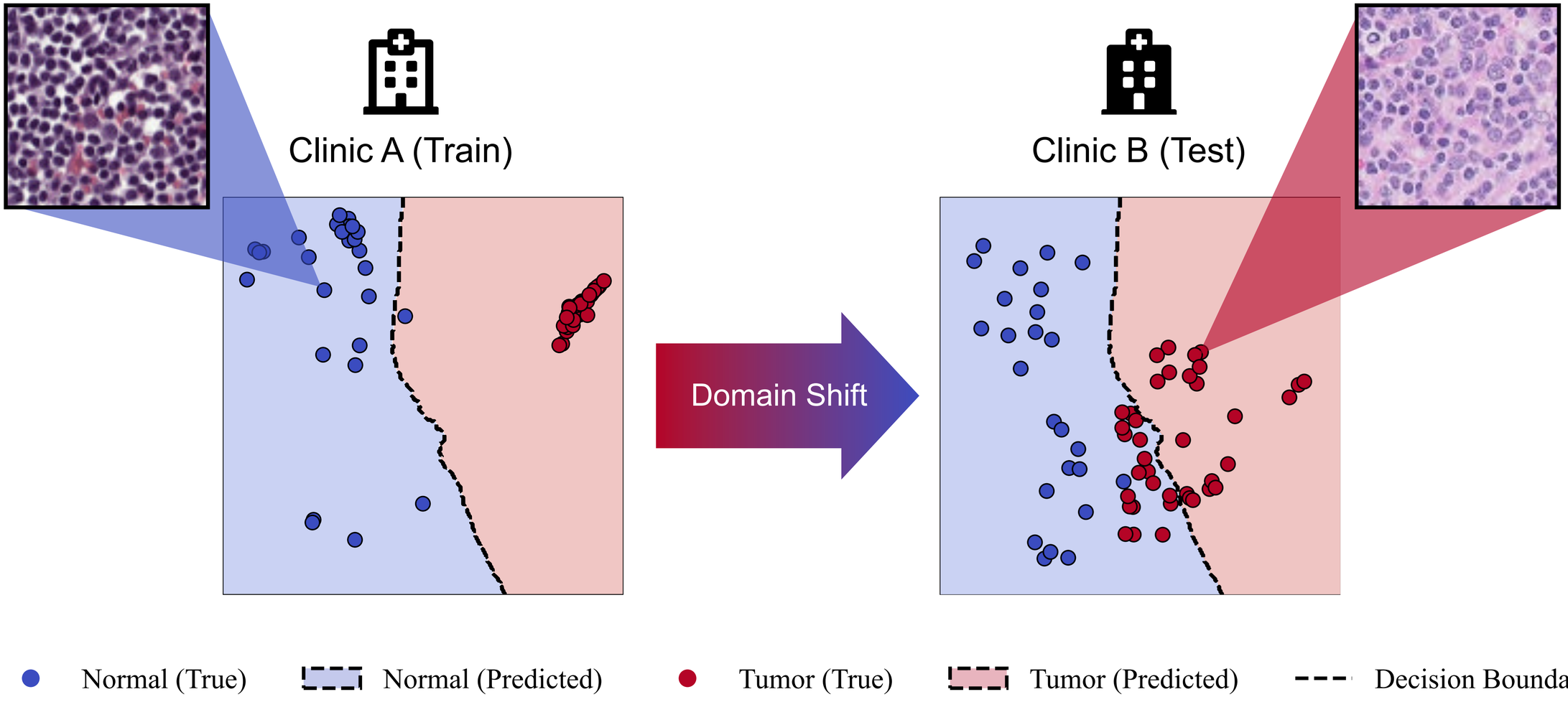
Feature Space Analysis



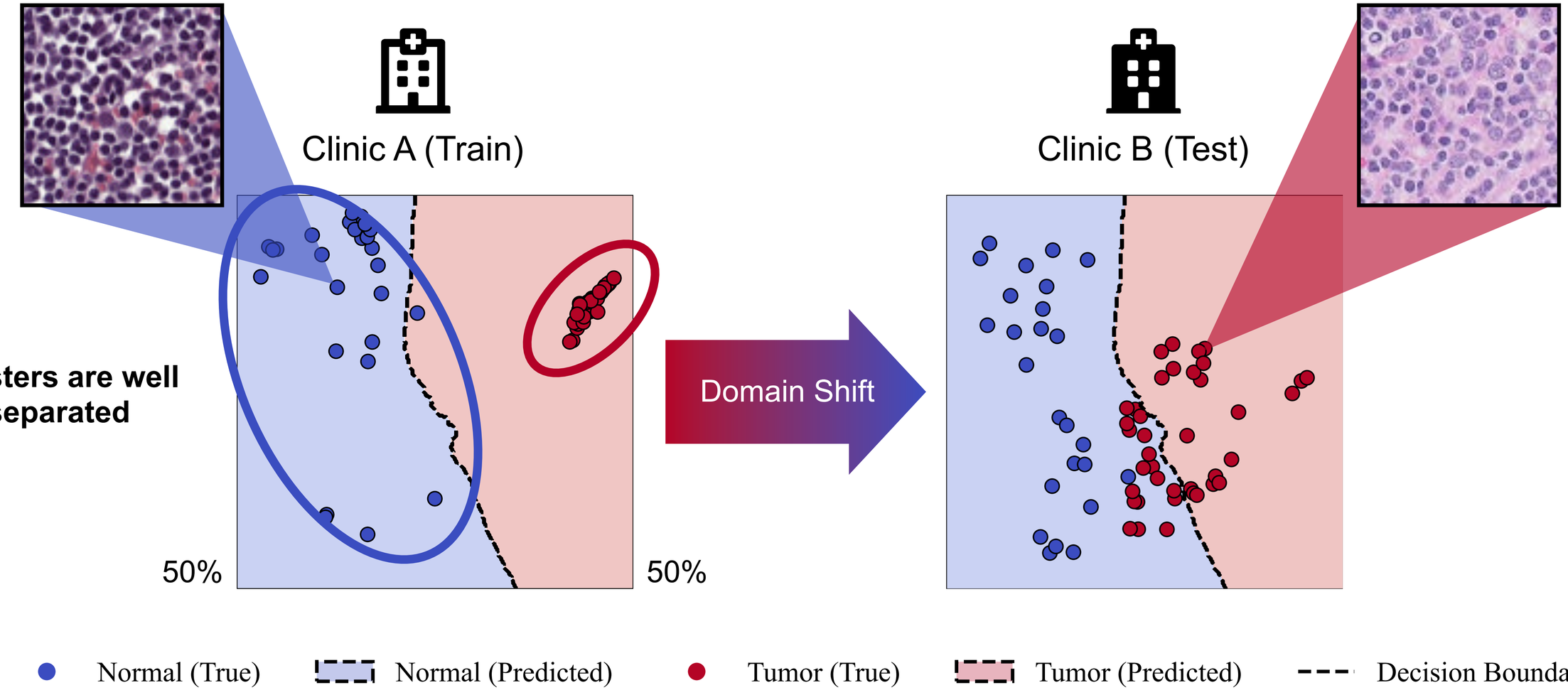
The classifier acts as a decision boundary in feature space
→ **stays fixed** during domain shift

Featurizer stays fixed during domain shift while the input distribution changes
→ embedded images also **move in feature space**

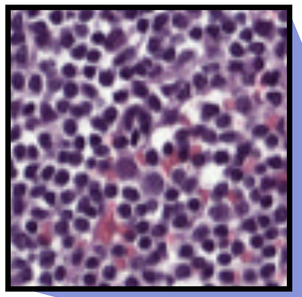
Model Collapse is Rooted Before EM Begins



Model Collapse is Rooted Before EM Begins



Model Collapse is Rooted Before EM Begins



Clinic A (Train)

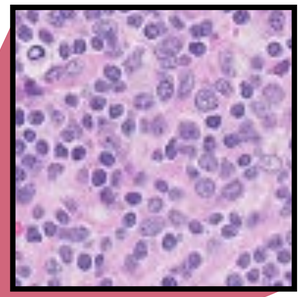
Clusters are well separated

50%

50%

Domain Shift

Clinic B (Test)



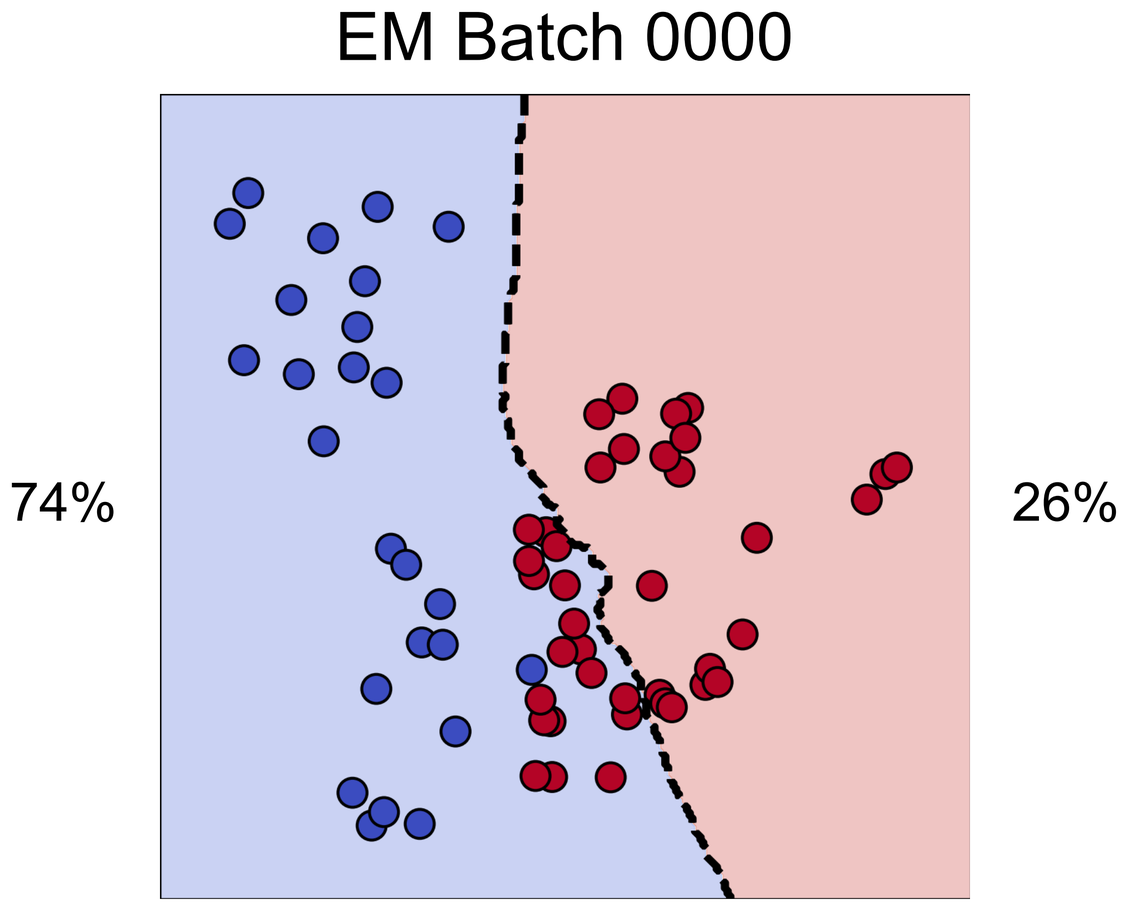
Clusters Merge → Inducing **Prediction Bias**

74%

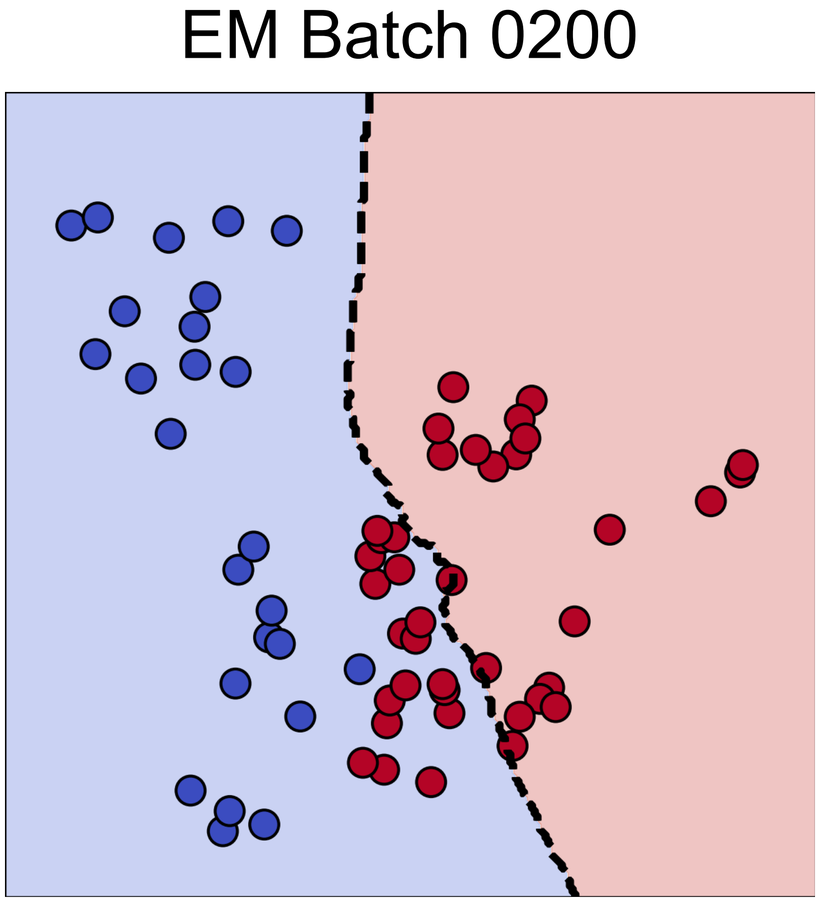
26%

Normal (True) Normal (Predicted) Tumor (True) Tumor (Predicted) Decision Boundary

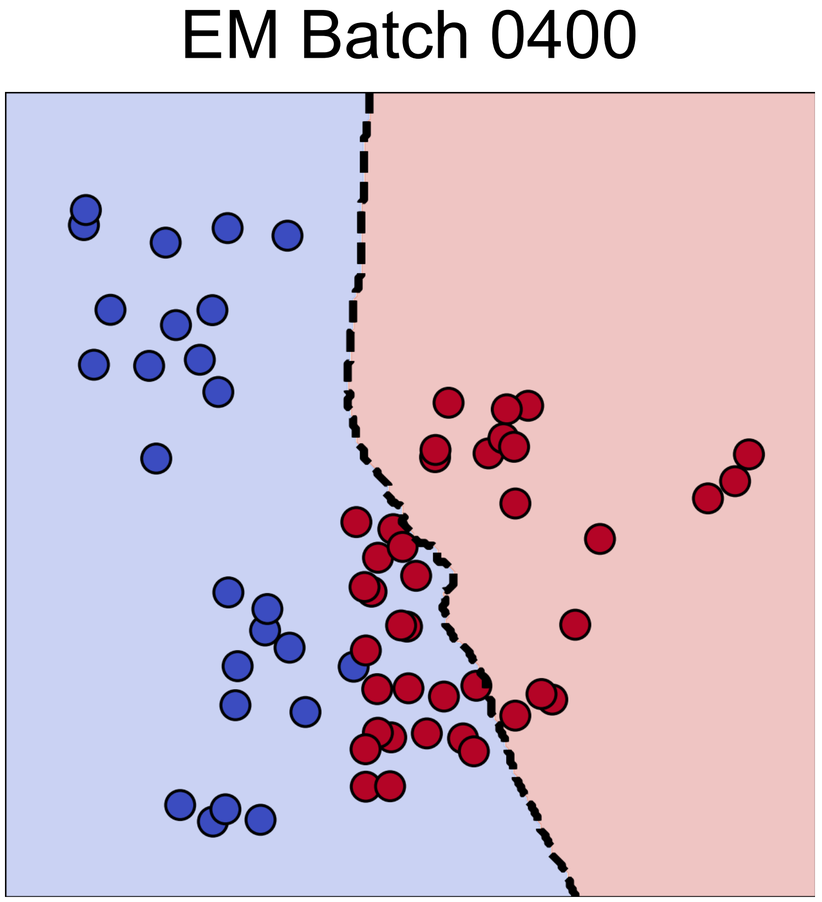
EM Amplifies Prediction Bias ↛ Model Collapse



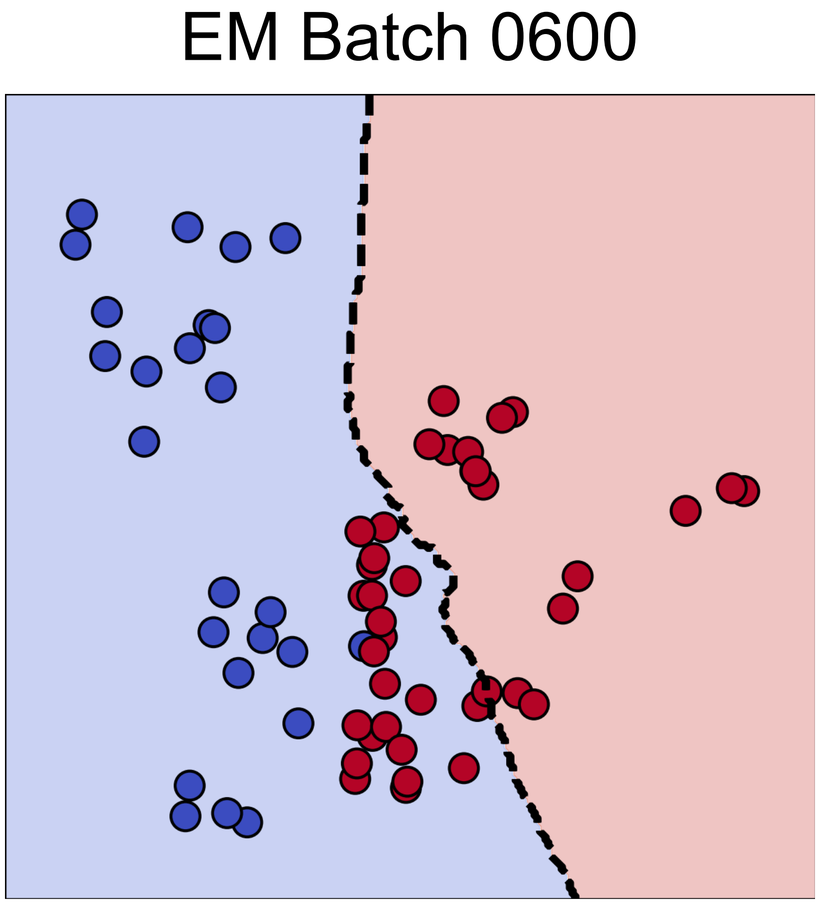
EM Amplifies Prediction Bias ↛ Model Collapse



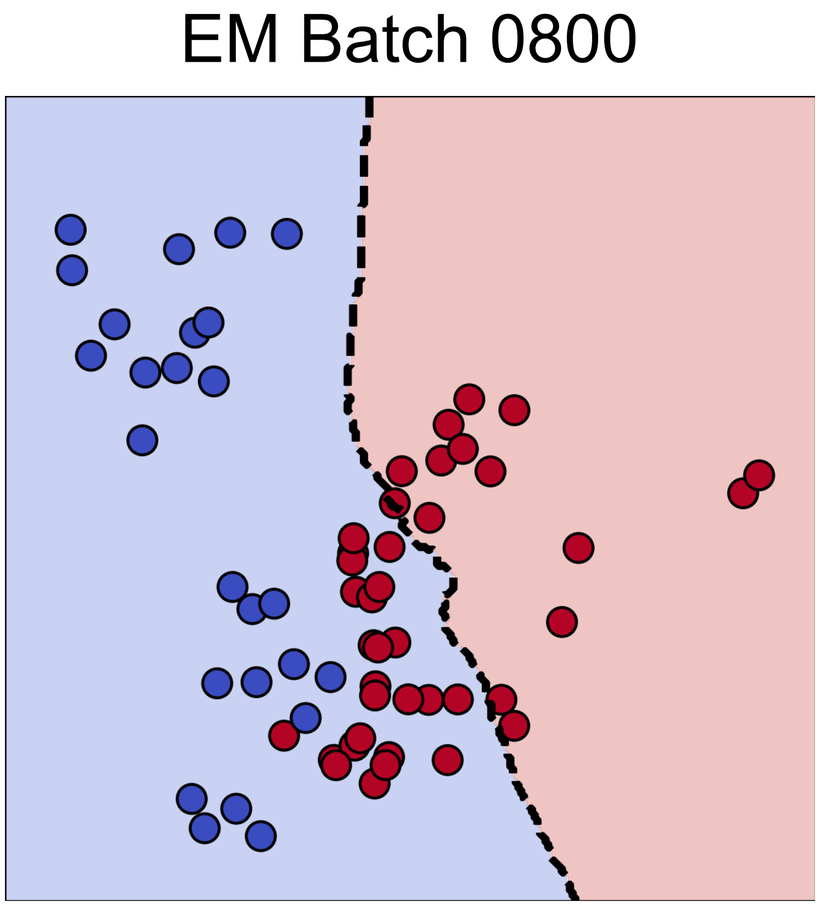
EM Amplifies Prediction Bias ↛ Model Collapse



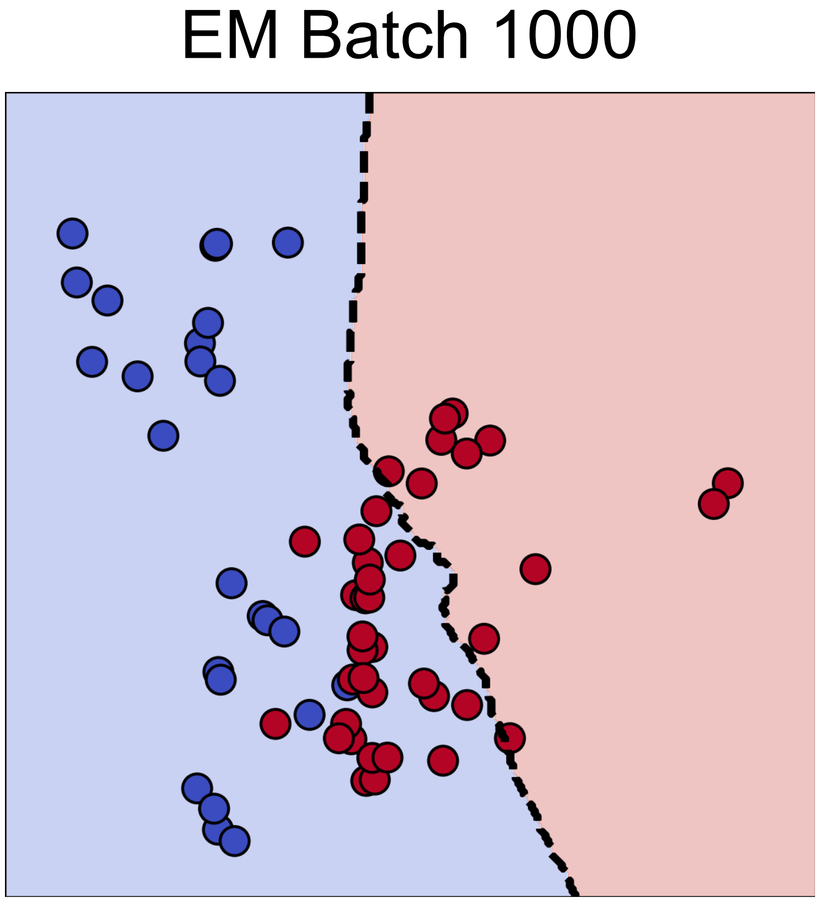
EM Amplifies Prediction Bias ↛ Model Collapse



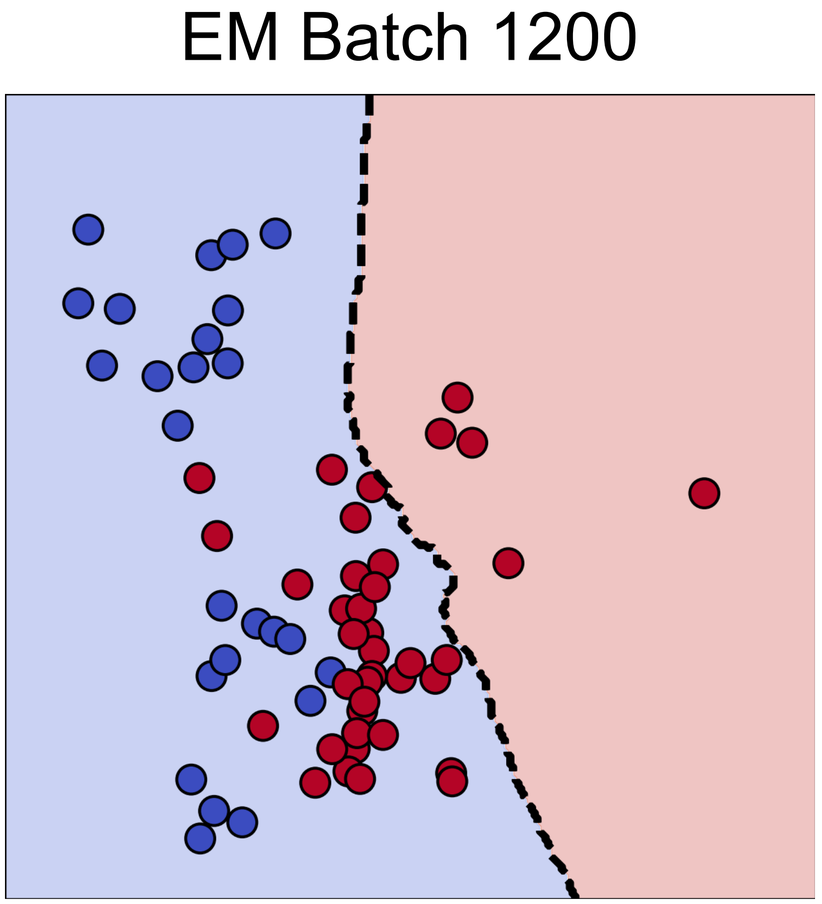
EM Amplifies Prediction Bias ↛ Model Collapse



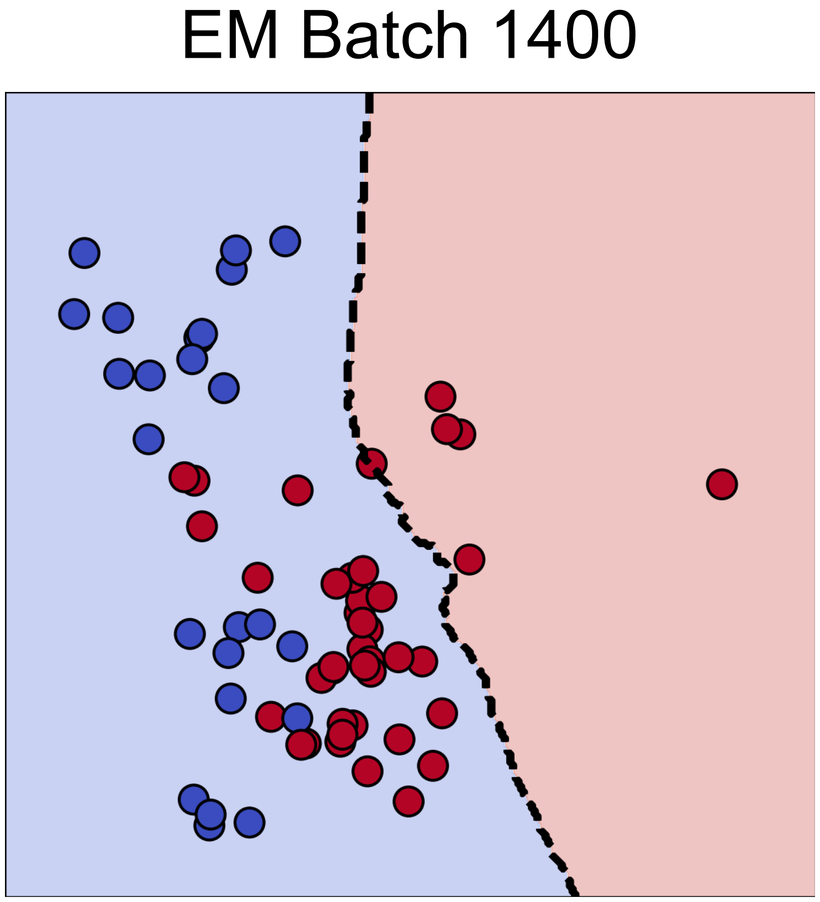
EM Amplifies Prediction Bias ↛ Model Collapse



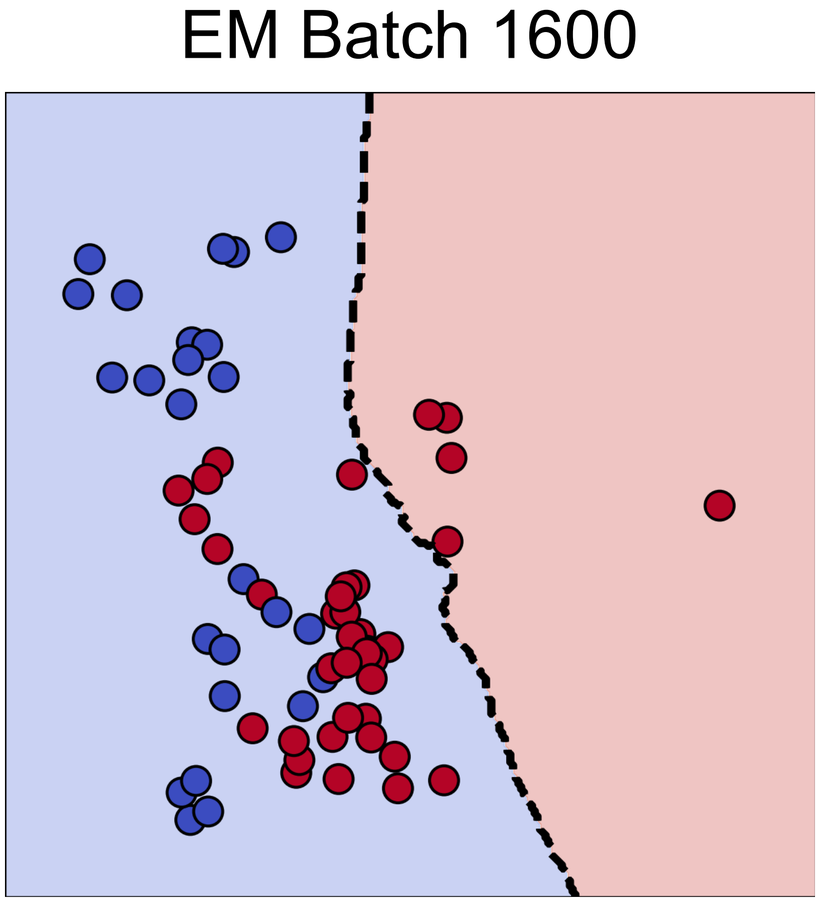
EM Amplifies Prediction Bias ↛ Model Collapse



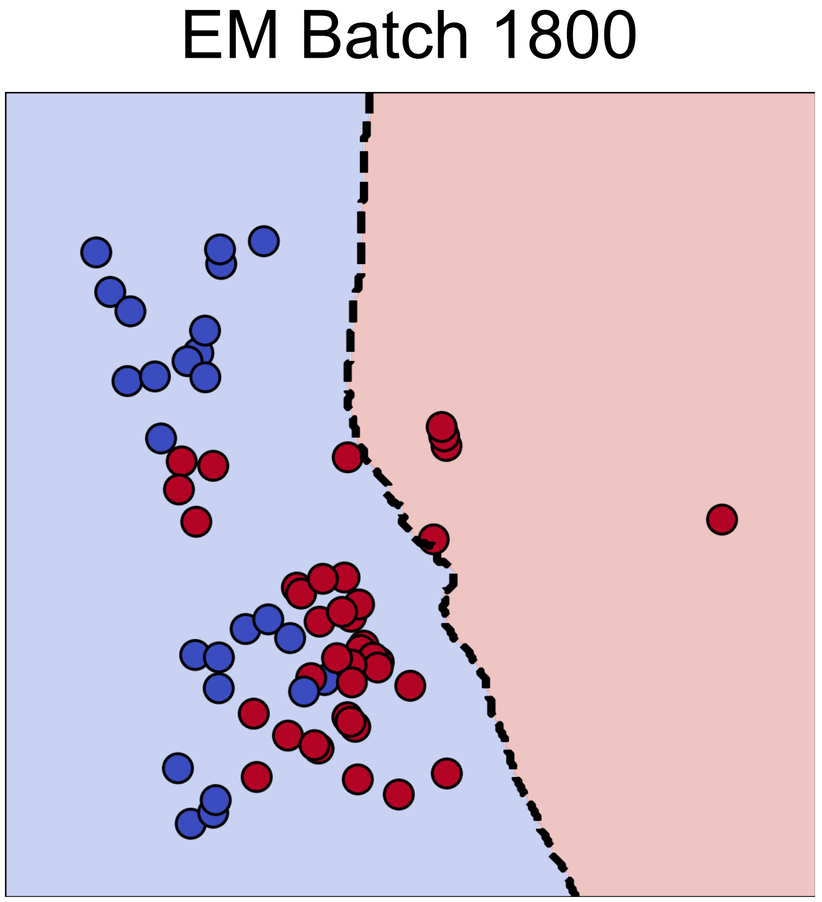
EM Amplifies Prediction Bias ↠ Model Collapse



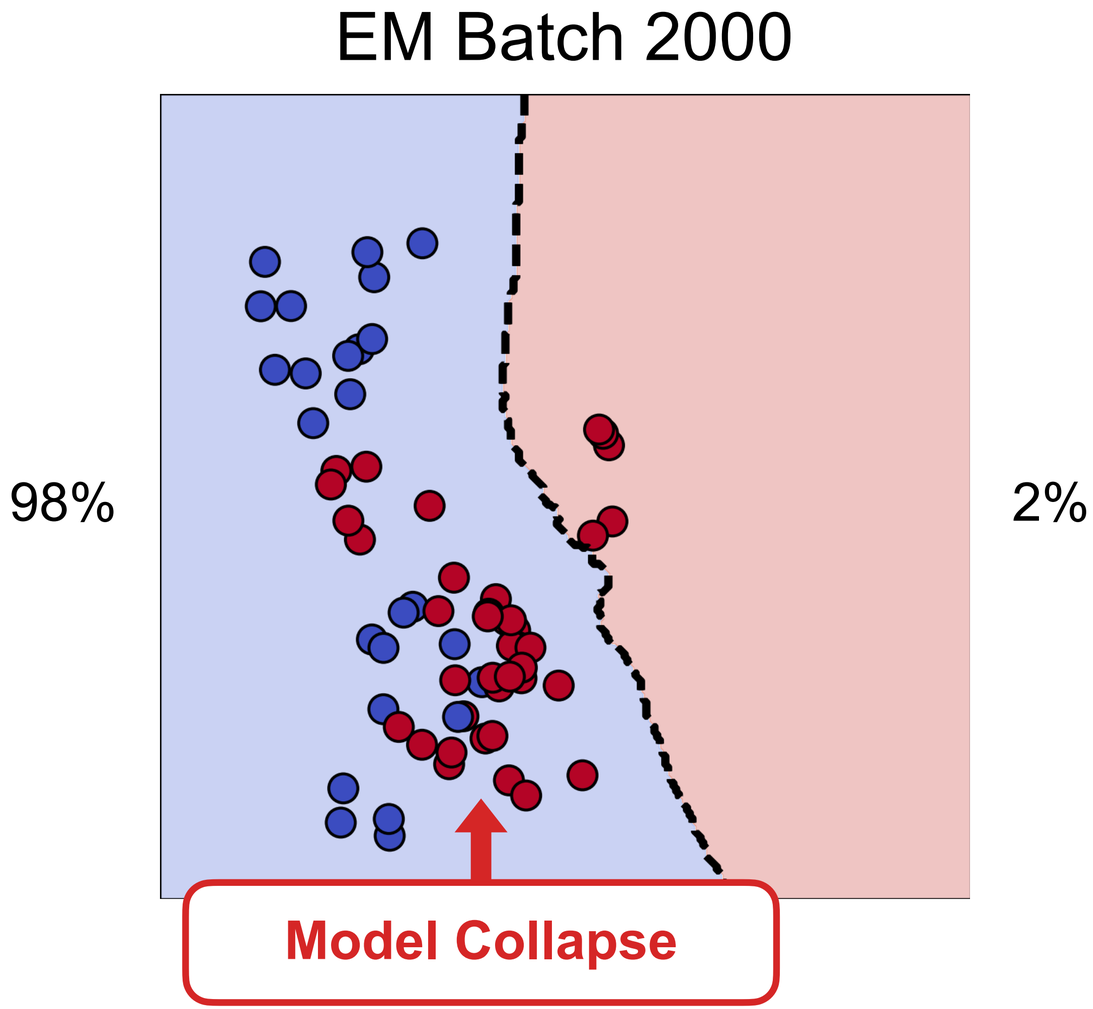
EM Amplifies Prediction Bias ↛ Model Collapse



EM Amplifies Prediction Bias ↛ Model Collapse



EM Amplifies Prediction Bias ↛ Model Collapse



Distribution Shift Bias Reduction (DSBR)

Importance Reweighting:

Given sample x, distributed according to variable $X \sim p(x)$, model outputs $f(x)$, and predictions $\hat{y}(x) = \arg\max_k f(x)_k$, the per sample weight is defined as :

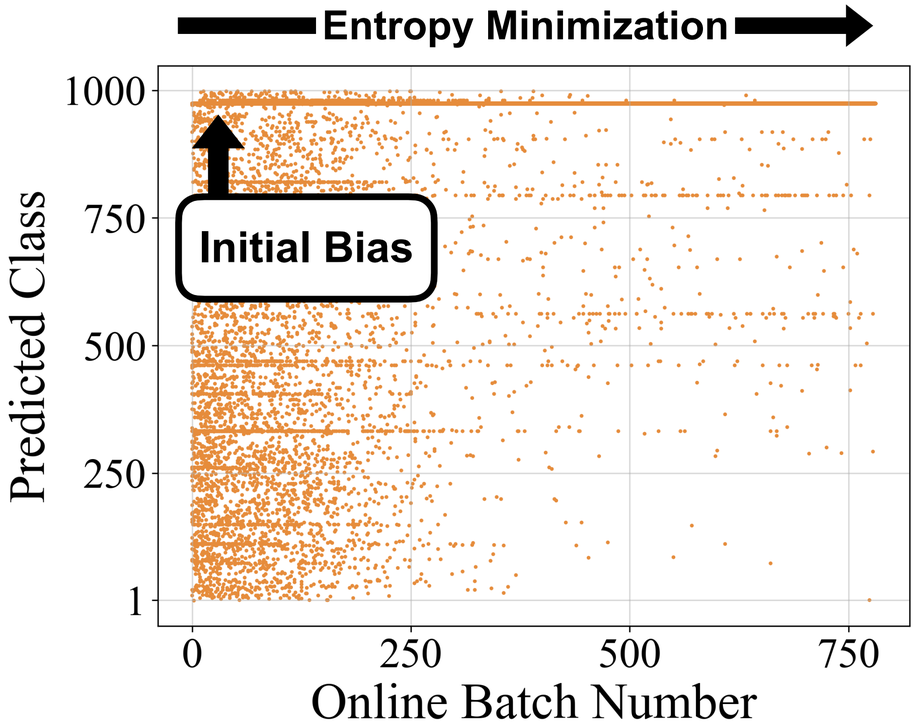
$$w(x) = \frac{1}{K} \cdot \mathrm{P}\left(\hat{y}(X) = \hat{y}(x)\right)^{-1}$$

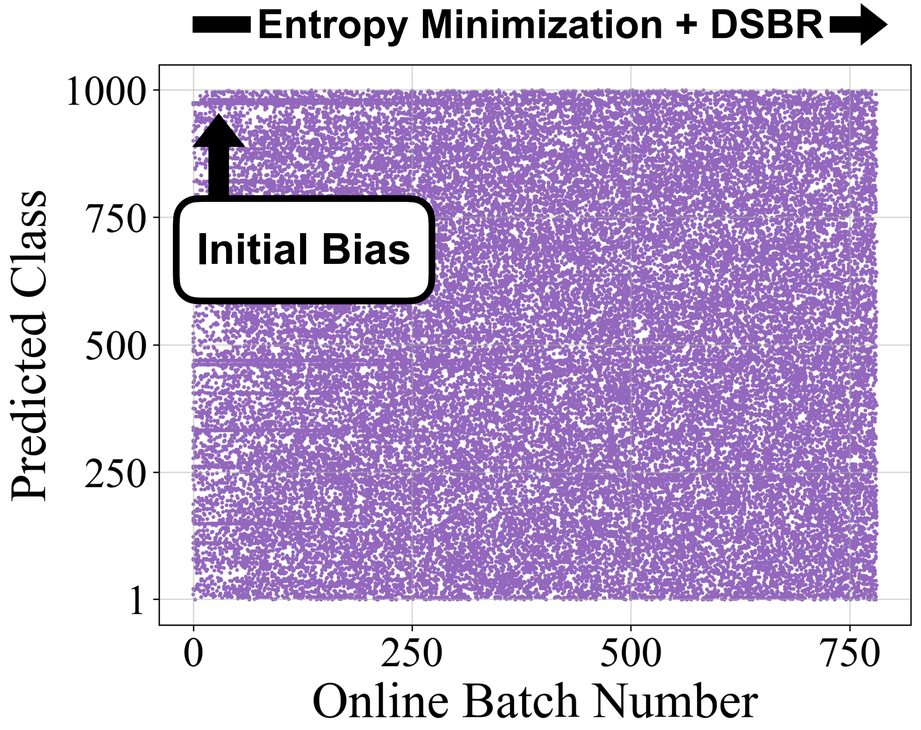
Balanced ÷ Predicted class probability

Key idea: Reweight samples such that every predicted class contributes equally to the loss.

Our DSBR Mitigates Prediction Bias

Model collapse: the model predicts the same class for all inputs

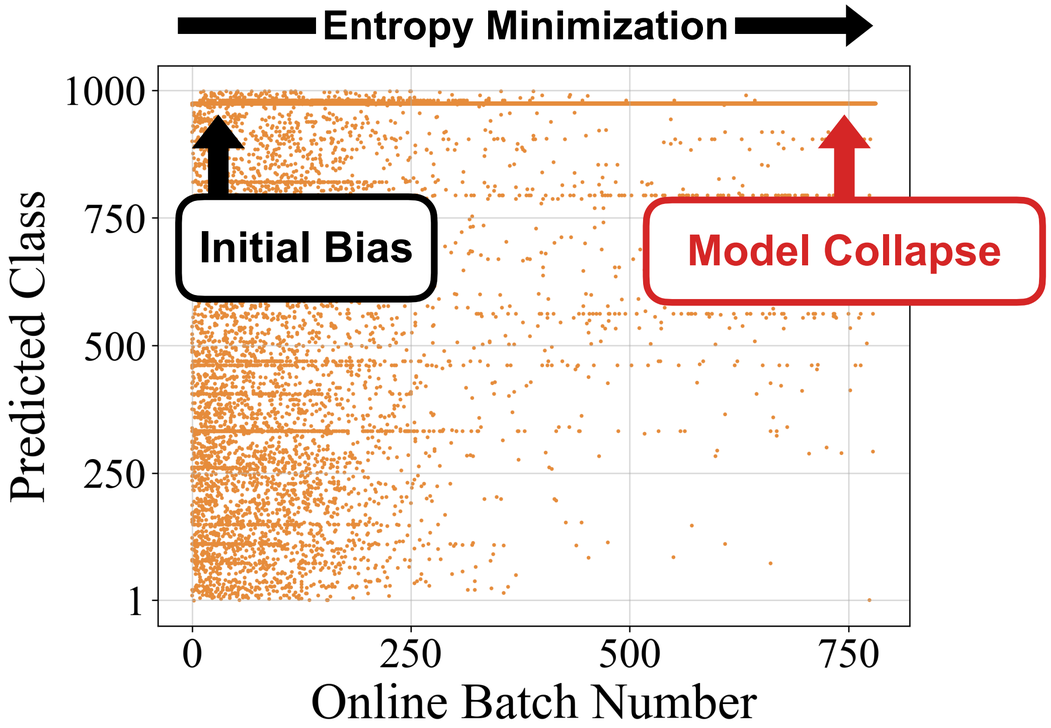


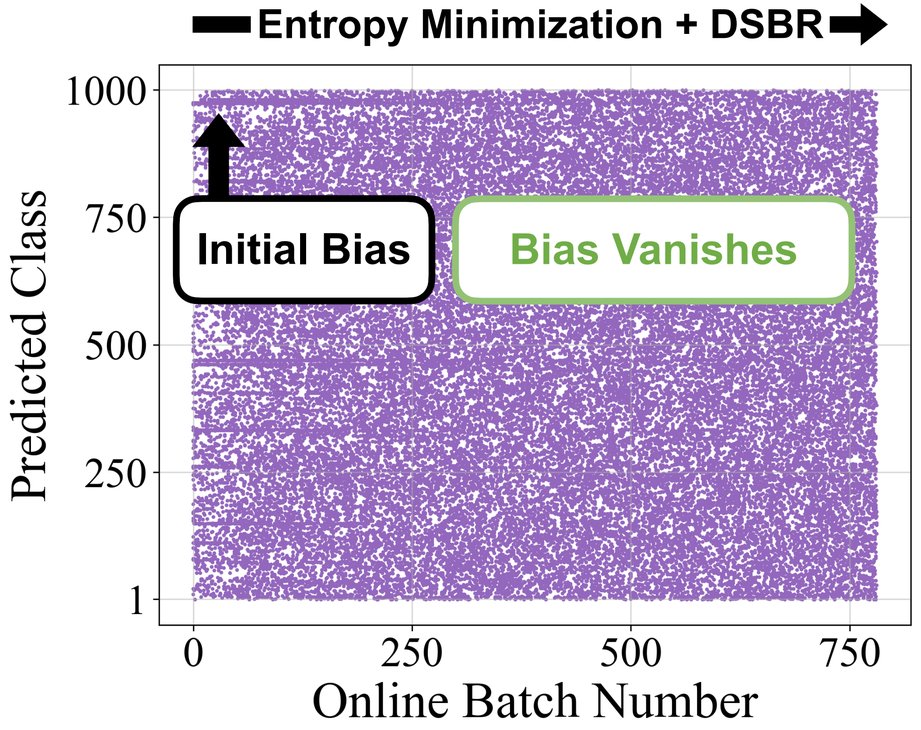


Prediction bias: models over-predict some classes and under-predict others.

Our DSBR Mitigates Prediction Bias

Model collapse: the model predicts the same class for all inputs





Prediction bias: models over-predict some classes and under-predict others.

DSBR succeeds on Medical Imaging Classification

Average balanced accuracy (%) across source-target configurations for four medical imaging datasets.

	Model & Method	**Histopantum**	**MammoBench**	**WILDSCamelyon**	**GliomaMRI**	**Δ Average**
Base Model	ResNet	$86.6_{\pm 0.0}$	$59.4_{\pm 0.0}$	$87.4_{\pm 0.0}$	$42.9_{\pm 0.0}$	▲ $0.0_{\pm 0.0}$
	• Tent [78]	$\underline{88.5}_{\pm 0.1}$	$58.8_{\pm 0.1}$	$87.9_{\pm 0.0}$	$\underline{44.3}_{\pm 1.3}$	▲ $0.8_{\pm 0.4}$
	• SAR [43]	$86.6_{\pm 0.0}$	$59.4_{\pm 0.0}$	$87.4_{\pm 0.0}$	$44.2_{\pm 1.4}$	▲ $0.3_{\pm 0.4}$
Ours	• ROID [51]	$87.2_{\pm 1.7}$	$56.6_{\pm 0.5}$	$\underline{94.8}_{\pm 0.6}$	$\mathbf{44.4}_{\pm 1.3}^{*}$	▲ $\underline{1.7}_{\pm 1.0}$
	• DeYO [44]	$76.1_{\pm 2.8}$	$\underline{60.9}_{\pm 0.2}$	$92.7_{\pm 0.5}$	n/a†	▼ $0.9_{\pm 0.9}$
	• ***DSBR*** (ours)	$\mathbf{89.9}_{\pm 0.4}$	$\mathbf{62.7}_{\pm 0.5}$	$\mathbf{96.5}_{\pm 0.1}$	$44.0_{\pm 1.1}$	▲ $\mathbf{4.2}_{\pm 0.5}$
Base Model	ViT	$87.1_{\pm 0.0}$	$54.4_{\pm 0.0}$	$87.0_{\pm 0.0}$	$37.2_{\pm 0.0}$	▲ $0.0_{\pm 0.0}$
	• Tent [78]	$87.4_{\pm 0.0}$	$54.4_{\pm 0.0}$	$87.4_{\pm 0.1}$	$\underline{37.4}_{\pm 0.2}$	▲ $0.2_{\pm 0.1}$
	• SAR [43]	$87.1_{\pm 0.0}$	$54.4_{\pm 0.0}$	$87.0_{\pm 0.0}$	$37.2_{\pm 0.0}$	▲ $0.0_{\pm 0.0}$
Ours	• ROID [51]	$\underline{88.0}_{\pm 0.3}$	$\underline{54.5}_{\pm 0.8}$	$\underline{94.3}_{\pm 0.2}$	$\mathbf{38.2}_{\pm 5.0}^{*}$	▲ $\underline{2.3}_{\pm 1.6}$
	• DeYO [44]	$87.9_{\pm 0.2}$	$\mathbf{55.2}_{\pm 0.1}$	$93.5_{\pm 0.3}$	n/a†	▲ $2.0_{\pm 0.2}$
	• ***DSBR*** (ours)	$\mathbf{88.8}_{\pm 0.4}$	$\mathbf{55.2}_{\pm 0.5}$	$\mathbf{95.6}_{\pm 0.3}$	$36.5_{\pm 0.9}$	▲ $\mathbf{2.6}_{\pm 0.5}$
		Tumor, Histology 🔬	Abnormality, X-ray 🩻	Tumor, Histology 🔬	Subtype, MRI 3D 🧠	

DSBR succeeds on Medical Imaging Classification

Average balanced accuracy (%) across source-target configurations for four medical imaging datasets.

	Model & Method	**Histopantum**	**MammoBench**	**WILDSCamelyon**	**GliomaMRI**	**Δ Average**
Base Model	ResNet	$86.6_{\pm 0.0}$	$59.4_{\pm 0.0}$	$87.4_{\pm 0.0}$	$42.9_{\pm 0.0}$	▲ $0.0_{\pm 0.0}$
Ours	• Tent [78]	$\underline{88.5}_{\pm 0.1}$	$58.8_{\pm 0.1}$	$87.9_{\pm 0.0}$	$\underline{44.3}_{\pm 1.3}$	▲ $0.8_{\pm 0.4}$
	• SAR [43]	$86.6_{\pm 0.0}$	$59.4_{\pm 0.0}$	$87.4_{\pm 0.0}$	$44.2_{\pm 1.4}$	▲ $0.3_{\pm 0.4}$
	• ROID [51]	$87.2_{\pm 1.7}$	$56.6_{\pm 0.5}$	$\underline{94.8}_{\pm 0.6}$	$\mathbf{44.4}_{\pm 1.3}^{*}$	▲ $\underline{1.7}_{\pm 1.0}$
	• DeYO [44]	$76.1_{\pm 2.8}$	$\underline{60.9}_{\pm 0.2}$	$92.7_{\pm 0.5}$	n/a†	▼ $0.9_{\pm 0.9}$
	• ***DSBR*** (ours)	$\mathbf{89.9}_{\pm 0.4}$	$\mathbf{62.7}_{\pm 0.5}$	$\mathbf{96.5}_{\pm 0.1}$	$44.0_{\pm 1.1}$	▲ $\mathbf{4.2}_{\pm 0.5}$
Base Model	ViT	$87.1_{\pm 0.0}$	$54.4_{\pm 0.0}$	$87.0_{\pm 0.0}$	$37.2_{\pm 0.0}$	▲ $0.0_{\pm 0.0}$
Ours	• Tent [78]	$87.4_{\pm 0.0}$	$54.4_{\pm 0.0}$	$87.4_{\pm 0.1}$	$\underline{37.4}_{\pm 0.2}$	▲ $0.2_{\pm 0.1}$
	• SAR [43]	$87.1_{\pm 0.0}$	$54.4_{\pm 0.0}$	$87.0_{\pm 0.0}$	$37.2_{\pm 0.0}$	▲ $0.0_{\pm 0.0}$
	• ROID [51]	$\underline{88.0}_{\pm 0.3}$	$\underline{54.5}_{\pm 0.8}$	$\underline{94.3}_{\pm 0.2}$	$\mathbf{38.2}_{\pm 5.0}^{*}$	▲ $\underline{2.3}_{\pm 1.6}$
	• DeYO [44]	$87.9_{\pm 0.2}$	$\mathbf{55.2}_{\pm 0.1}$	$93.5_{\pm 0.3}$	n/a†	▲ $2.0_{\pm 0.2}$
	• ***DSBR*** (ours)	$\mathbf{88.8}_{\pm 0.4}$	$\mathbf{55.2}_{\pm 0.5}$	$\mathbf{95.6}_{\pm 0.3}$	$36.5_{\pm 0.9}$	▲ $\mathbf{2.6}_{\pm 0.5}$
		Tumor, Histology 🔬	Abnormality, X-ray 🩻	Tumor, Histology 🔬	Subtype, MRI 3D 🧠	

Moreover, DSBR also stays robust in natual vision's ImageNet-C benchmark for both ViT and ResNet

Key Takeaways

1. Model collapse is rooted **before** EM begins (domain shift introduces **prediction bias**)

2. EM **Amplifies** this during unsupervised learning

3. Bias reduction using DSBR is a simple but effective strategy to preventing model collapse across multiple modalities