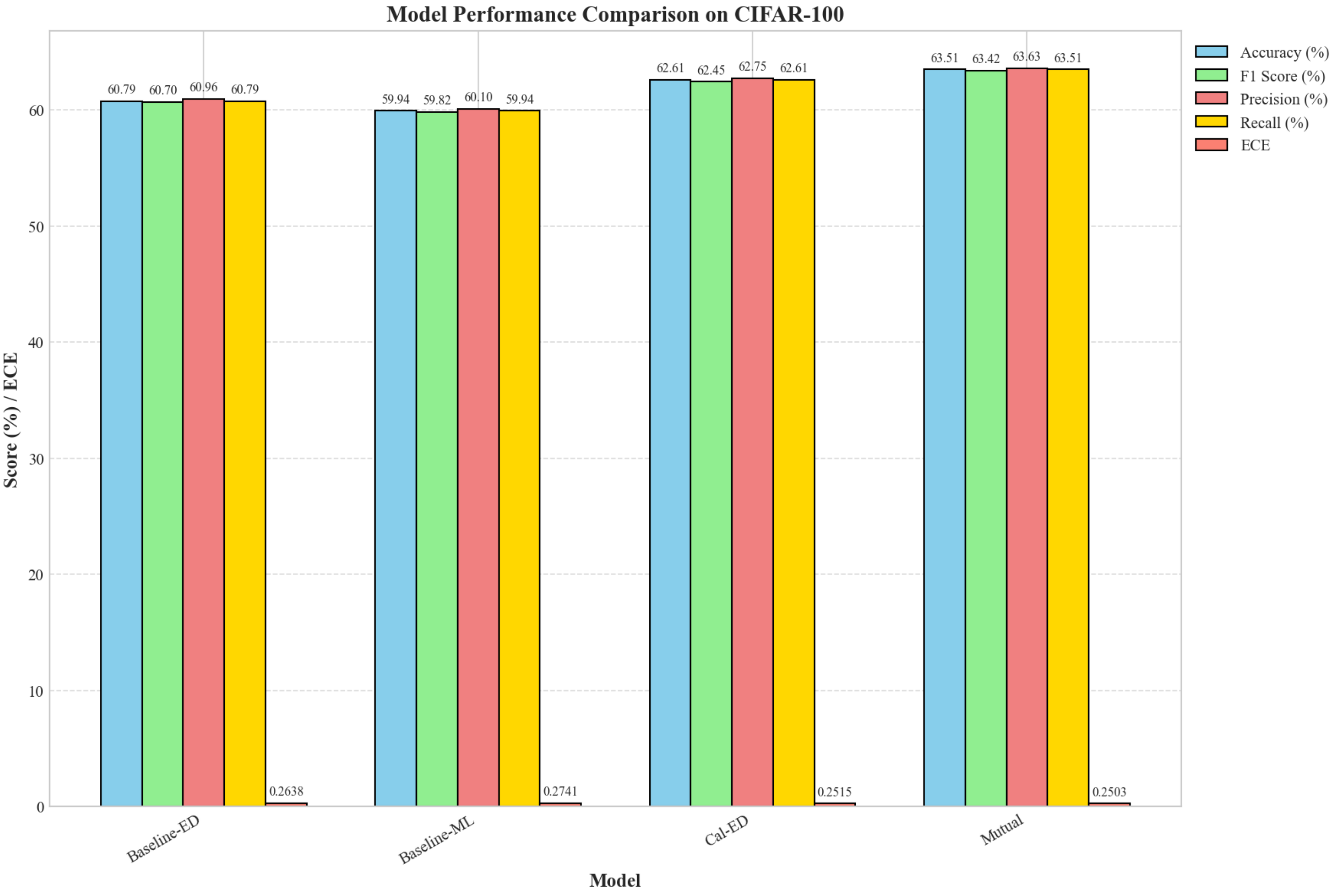
CALM: CIFAR-100 Fine-tuning Results (Test Set)

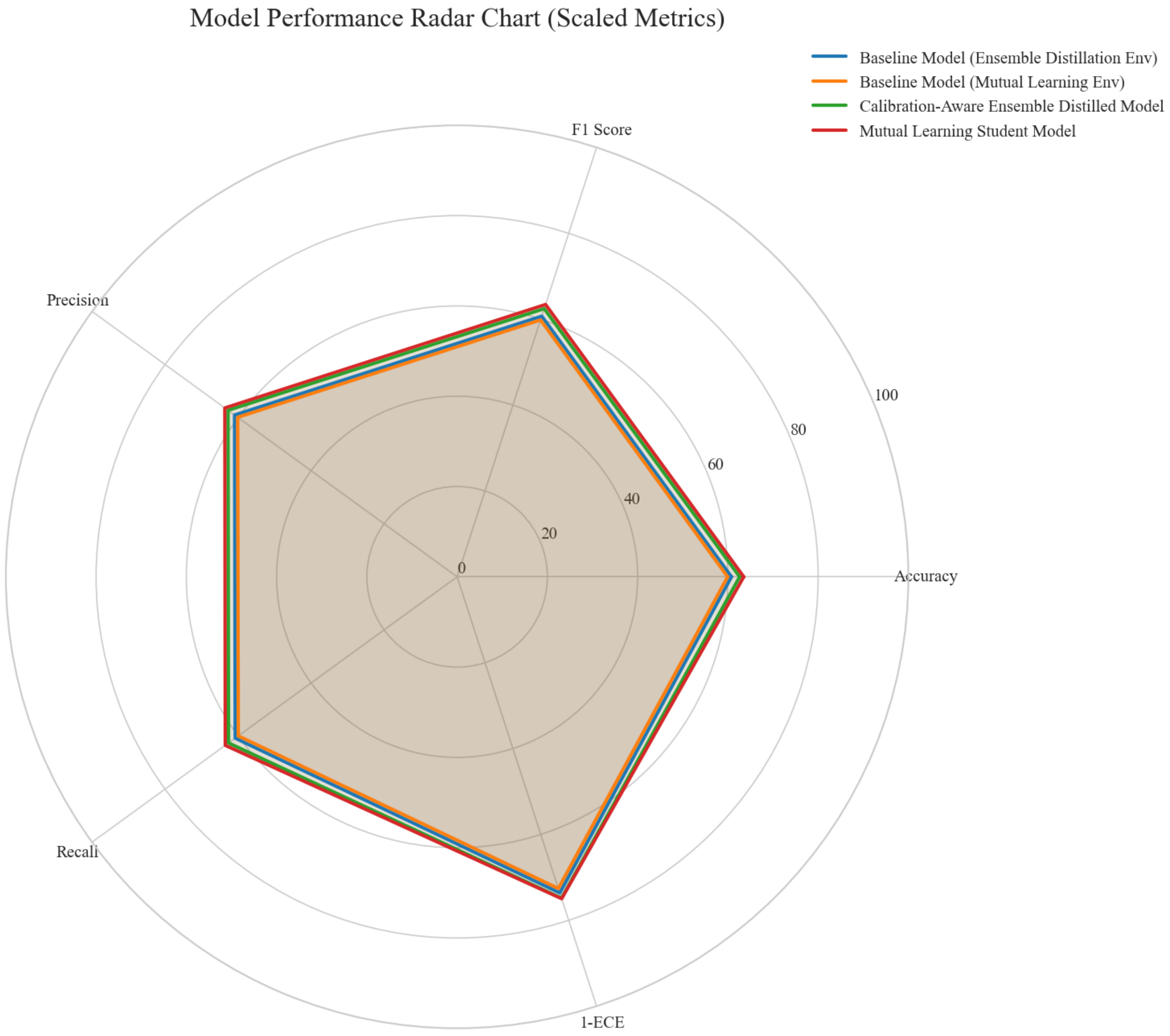
Model Performance Summary Table (Test Set)

Model	Acc (%)	F1 (%)	Prec (%)	Recall (%)	ECE
Baseline-ED	60.79	60.70	60.96	60.79	0.2638
Baseline-ML	59.94	59.82	60.10	59.94	0.2741
Cal-ED	62.61	62.45	62.75	62.61	0.2515
Mutual	63.51	63.42	63.63	63.51	0.2503

Performance Comparison (Bar Chart - Test Set)

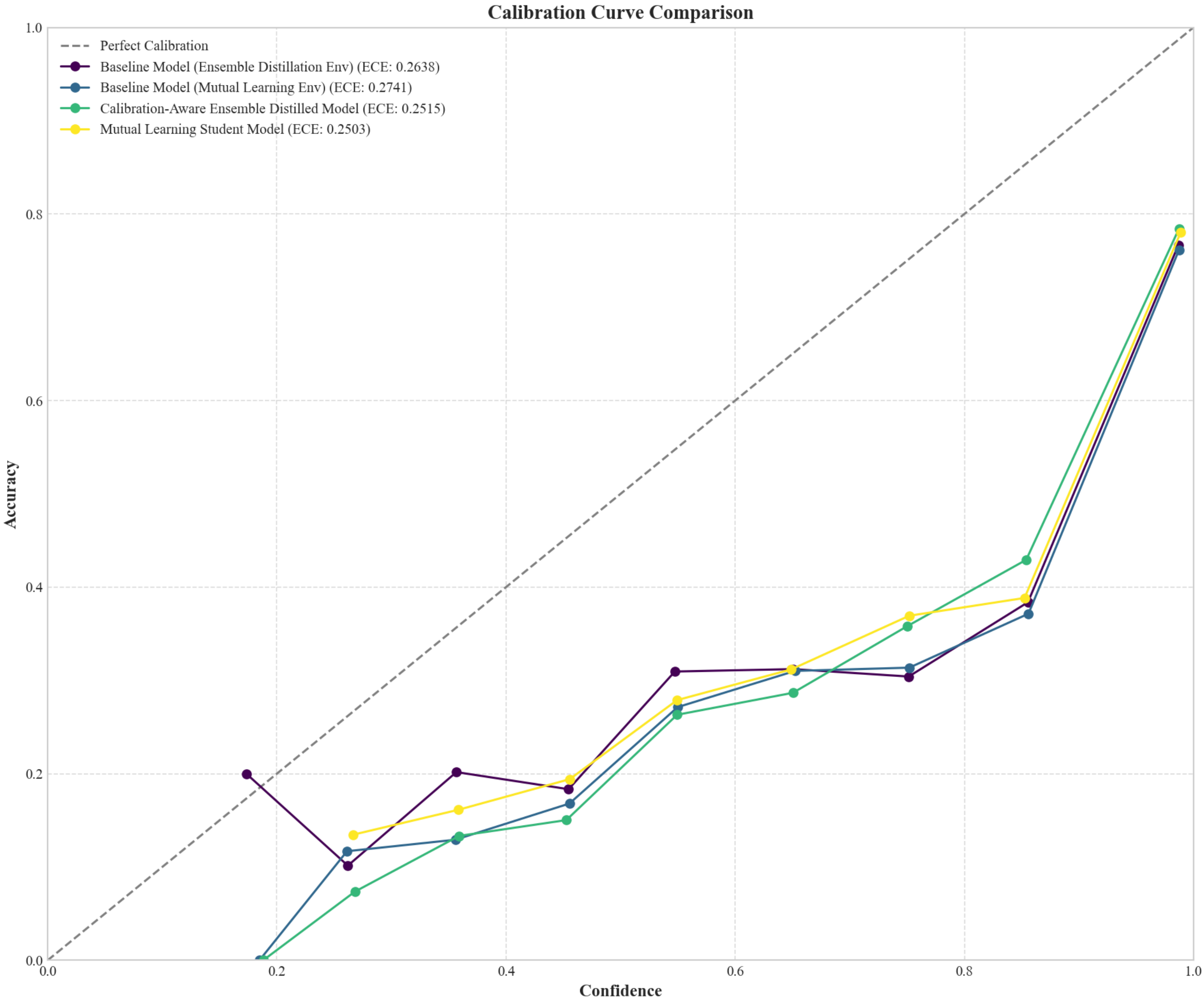


Performance Comparison (Radar Chart - Test Set)



CALM: CIFAR-100 Fine-tuning Results (Test Set)

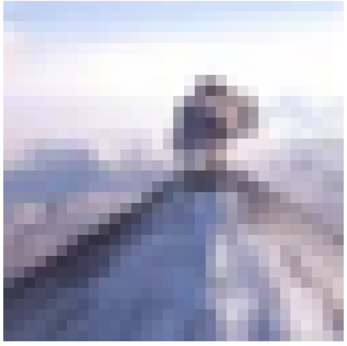
Calibration Curve Comparison (Test Set)



CALM: CIFAR-100 Fine-tuning Results (Test Set)

GradCAM Visualization Example 1

Example 1 - True Class: mountain



GradCAM Failed for Baseline Model (Ensemble Distillation Env)

GradCAM Failed for Baseline Model (Mutual Learning Env)

GradCAM Failed for Calibration-Aware Ensemble Distilled Model

GradCAM Failed for Mutual Learning Student Model

CALM: CIFAR-100 Fine-tuning Results (Test Set)

GradCAM Visualization Example 2

Example 2 - True Class: forest

GradCAM Failed for Baseline Model (Ensemble Distillation Env)

GradCAM Failed for Baseline Model (Mutual Learning Env)

GradCAM Failed for Calibration-Aware Ensemble Distilled Model

GradCAM Failed for Mutual Learning Student Model

CALM: CIFAR-100 Fine-tuning Results (Test Set)

GradCAM Visualization Example 3

Example 3 - True Class: seal

GradCAM Failed for Baseline Model (Ensemble Distillation Env)

GradCAM Failed for Baseline Model (Mutual Learning Env)

GradCAM Failed for Calibration-Aware Ensemble Distilled Model

GradCAM Failed for Mutual Learning Student Model