Figure 23. (**Visualizing MD tree and its diagnosis results on models test on out-of-distribution data (CIFAR-10-C)**). *Left side*: structure of the MD tree. The color of the leaf node indicates the predicted class by `MD tree`. The threshold values are learned from the training set. *Right side*: The first row represents training samples, and the second row represents test samples. Each colored circle represents one sample (which is one pre-trained model), and the color represents the ground-truth label: blue means the hyperparameter is too large, while red means small. The black dashed line indicates the decision boundary of `MD tree`. Each numbered regime on the right corresponds to the leaf node with the same number on the tree. The samples in 23(b) and those in 23(c) are separated by training error. The same applies to 23(d) and 23(e).
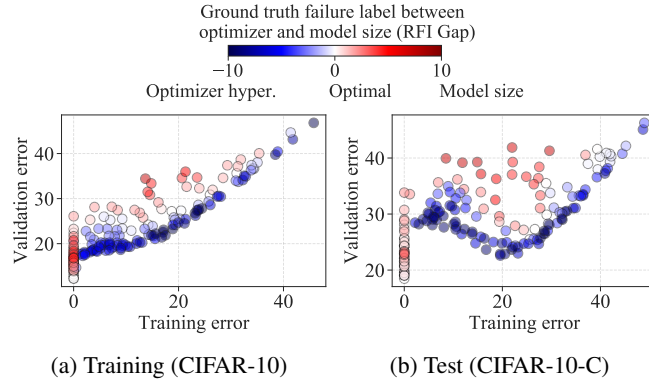

Figure 24. (**Applying validation metrics to diagnose the models tested on out-of-distribution data (CIFAR-10-C)**). (a) models in the training set are evaluated on ID CIFAR-10 test data, (b) models in the test set are evaluated on OOD CIFAR-10-C test data.

8