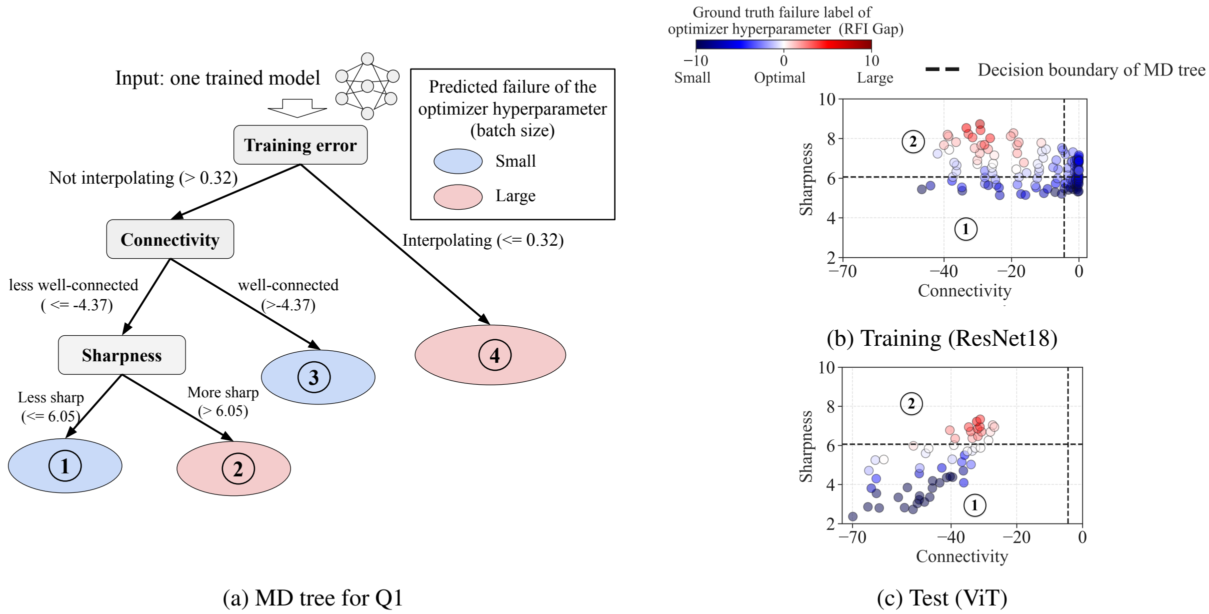

Figure 29. **(Visualizing MD tree and its diagnosis results on ViT-tiny models trained on CIFAR-10).** *Left side*: structure of the MD tree. The color of the leaf node indicates the predicted class by `MD tree`. The threshold values are learned from the training set. *Right side*: The first row represents 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.

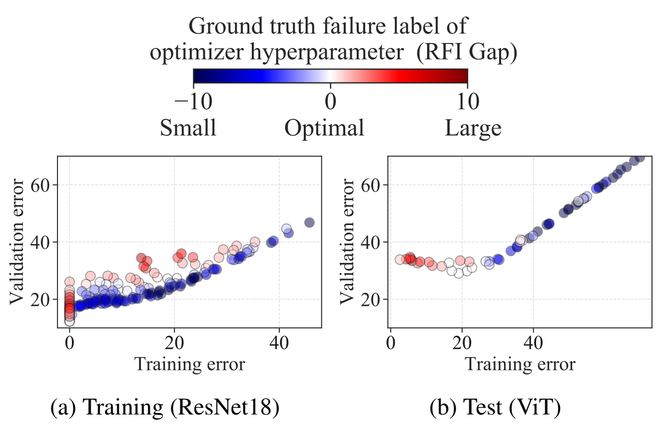

Figure 30. **(Applying validation metrics to diagnose the ViT-tiny models trained with CIFAR-10).** (a) training set comprises ResNet18 models trained on CIFAR-10, (b) test set comprises ViT-tiny models trained on CIFAR-10.

Figure 29 partially explains why the MD tree performs well in this case. We find that the ViT models mostly belong to Regimes 1 and 2 in our MD tree built on ResNet18. This indicates that Regimes 3 and 4 do not exist for these ViT models. Therefore, we indeed observe the difference of loss landscape properties between convolutional neural networks and transformer architectures. However, we still observe that the decision boundary of the sharpness metric that separates Regime 1 and 2 can perfectly transfer from ResNet18 models to ViT models. Therefore, the MD tree achieves good performance in diagnosing the ViT models.

Figure 30 shows that there exists a significant shift in the training/validation error range between ResNet18 models in the training set and the ViT-tiny model in the test set. Therefore, transferring diagnostic results from ResNet18 models to ViT models based on validation features is challenging.