



Figure 22. (**Diagnosing trained models tested on out-of-distribution data (CIFAR-10-C)**). The  $y$ -axis indicates the diagnosis accuracy. (a) The  $x$ -axis indicates the number of pre-trained models used for building the training set. (b) The  $x$ -axis indicates the maximum amount of training (image) data for training models in the training set. (c) The  $x$ -axis indicates the maximum number of parameters of the models in the training set. Our method MD tree outperforms two baseline methods in dataset transfer and scale transfer scenarios.