



Figure 21. (Comparing MD tree to additional baseline method (Hyperparameter + Validation + DT) on Q1 tasks of predicting optimizer hyperparameter is large or small). 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. The additional baseline outperforms the other two baseline methods, but our method MD tree still outperforms all three in dataset transfer and scale transfer scenarios.