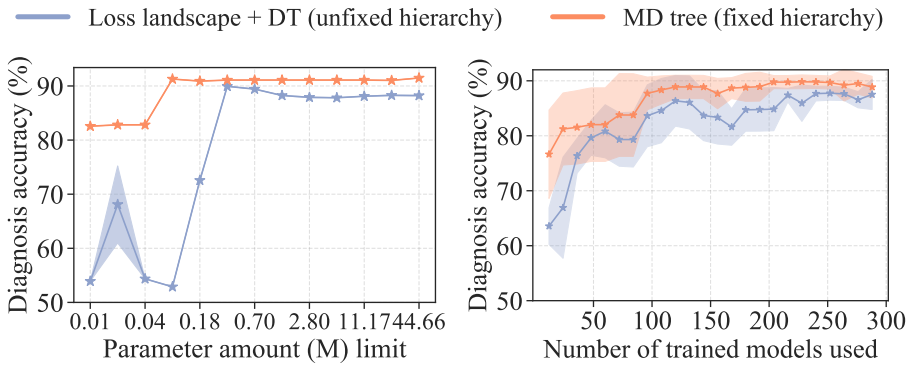

(a) Scale transfer (b) Dataset transfer

Figure 20. **(The fixed hierarchy of MD tree helps `MD tree` to generalize to unseen models with larger scales).** The y-axis indicates the diagnosis accuracy. (a) The x-axis indicates the maximum number of parameters of the models in the training set. (b) The x-axis indicates the number of trained models used. `MD tree` outperforms normal decision tree with unfixed hierarchy trained with loss landscape metrics.

We provide new experiments to demonstrate the importance of the fixed tree hierarchy is important for the generalization ability of the MD tree in scale transfer and dataset transfer. We study the Q1 task of predicting whether the optimizer hyperparameter of a given model configuration is large or small. We study two methods: 1) MD tree is our method, which uses a fixed hierarchy. 2) "Loss landscape + DT" is the baseline method, which uses the same set of loss landscape features as the MD tree, but uses an unfixed and learnable architecture.

The results are shown in Figure 20. Figure 20(a) shows that the MD tree achieves superior performance (82.56% - 91.28%) compared to the baseline method (53.9% - 72.56%) when trained on small-scale models (less than 0.18M parameters). The advantage of the MD tree holds for the dataset transfer in Figure 20(b). Therefore, both results emphasize the importance of fixed tree hierarchy in improving the generalizability of the diagnosis method.