

Figure 1: The average test rank of all methods on the considered benchmark.

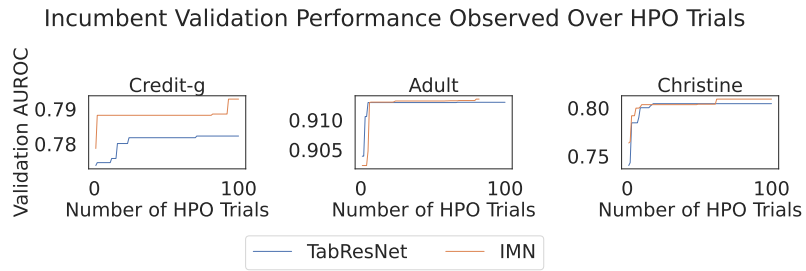


Figure 2: The incumbent validation performance during HPO.

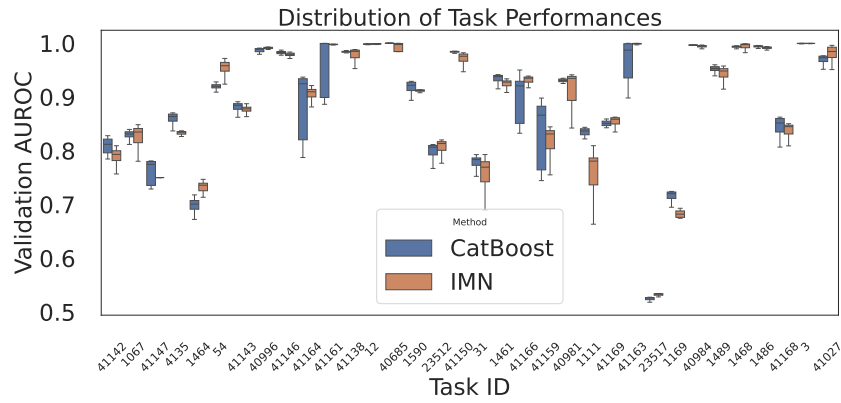


Figure 3: The distribution of the validation performance of the different hyperparameter configurations per task for CatBoost and IMN.