

Additional file 16: Fig. S6: Details of individual cross-validation runs of Fig. 2 on the performance of different random subsets containing 10%, 20%, 40%, 60%, 80%, and 100% of the dataset.

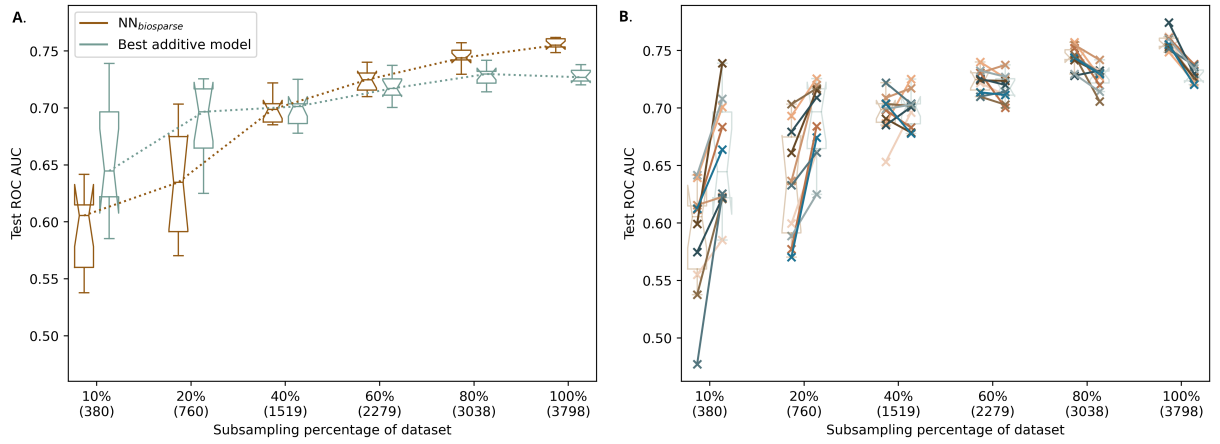


Figure 1: (A) Performance using different random subsets containing 10%, 20%, 40%, 60%, 80% and 100% of the dataset. Models shown are the best additive model (L_2 penalty) and $NN_{\text{biosparse}}$. Performances are measured using ten different threefold cross-validation runs, using identical splits for both models. (B) Same performance plot with lines connecting corresponding individual cross-validation runs with identical splits between the best linear model and the best neural network. Different colors represent different cross-validation run.