Performance scores calculated relative to mean and SD across all sets for each problem. Columns organised by descending overall mean balanced accuracy.

Rows organised by hierarchical clustering.

