

Head to head of feature sets over maximum of 102 problems

t-tests between accuracy distributions for each feature set and problem combination were calculated

