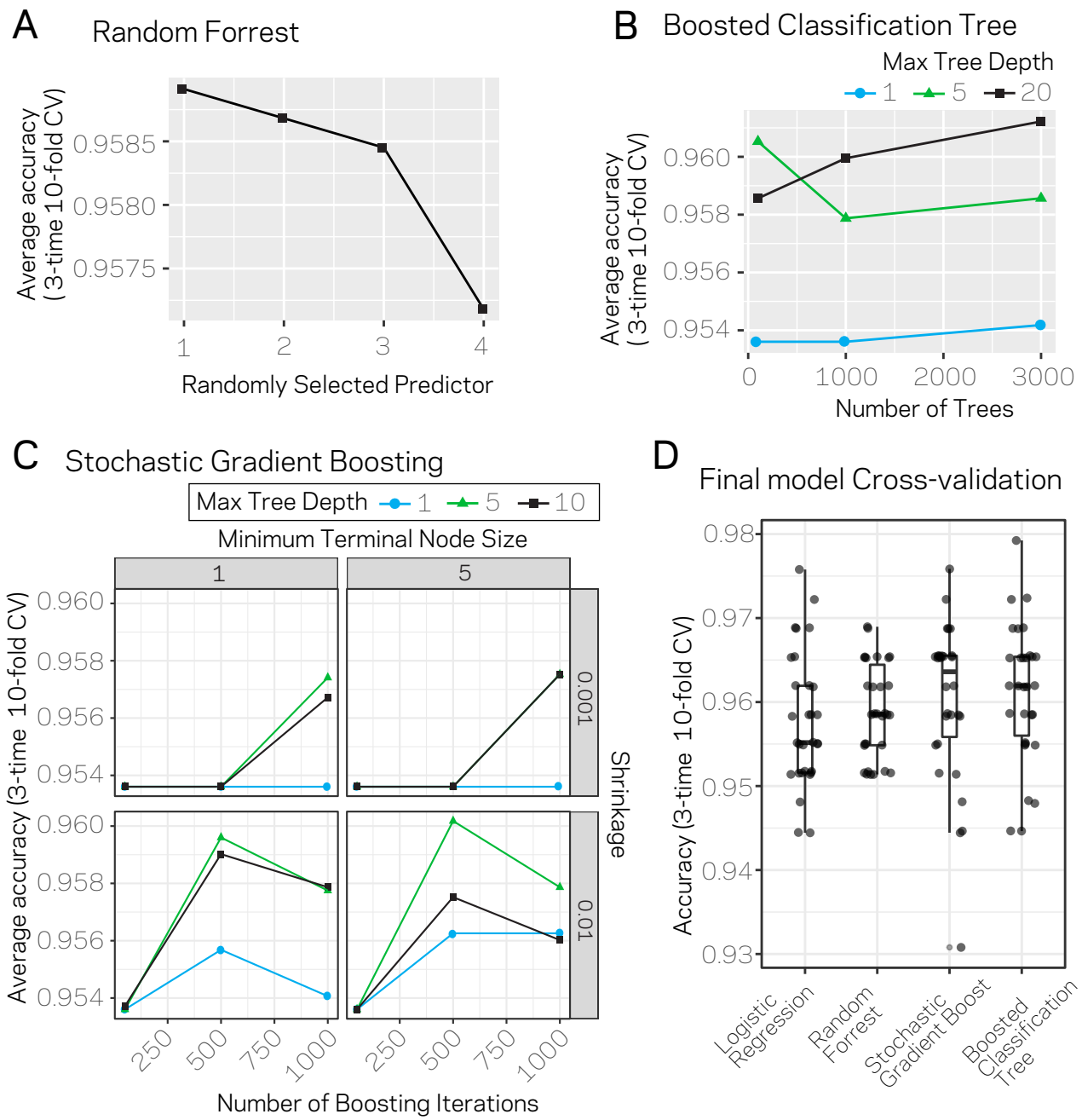

Figure 5- figure supplement 1. Results from model parameter tuning and assessment. (A) The random forrest model was trained using 3-time 10-fold cross-validation (CV) under varying conditions for the parameter 'randomly selected predictor'. The plot shows the average accuracy across the 30 trials for each variable level as a black square. (B) Average accuracy results from 3-time 10-fold CV using the boosted classification tree algorithm. The variables 'number of trees' (x-axis) and 'max tree depth' (blue, green, black lines) were varied across the trials. Each data point represents the average accuracy across the CV. (C) Average accuracy results from 3-time 10-fold CV using the stochastic gradient boosting algorithm. The variables 'number of boosting iterations' (x-axis), 'shrinkage' (y-axis), 'minimum terminal node size' (columns), and max tree depth (blue, green, black lines) were varied across the trials. Each data point represents the average accuracy across the CV. (D) The individual accuracy measurements and box plot for the final models picked for each algorithm. Results are from the 30 CV runs.