



Figure S1. Visualization of the 100x stratified random-split cross-validation, including a second cross-validation within the training set to perform the automatic workflows optimization. Optimization was done solely on the training set in order to prevent overfitting on the test set. The ensemble averages the predictions of the best 50 performing workflows to create a more robust model.