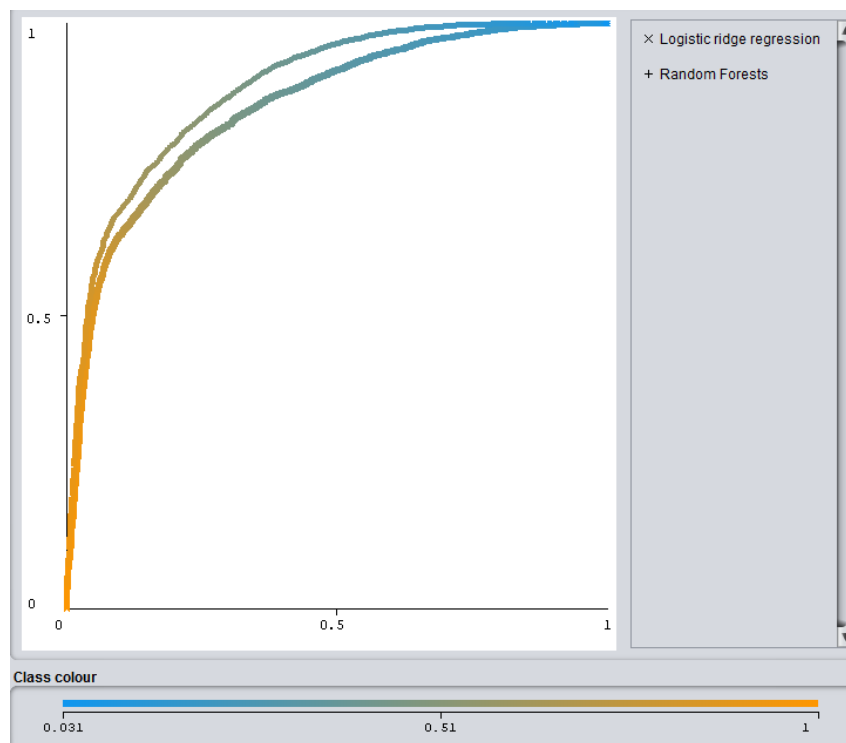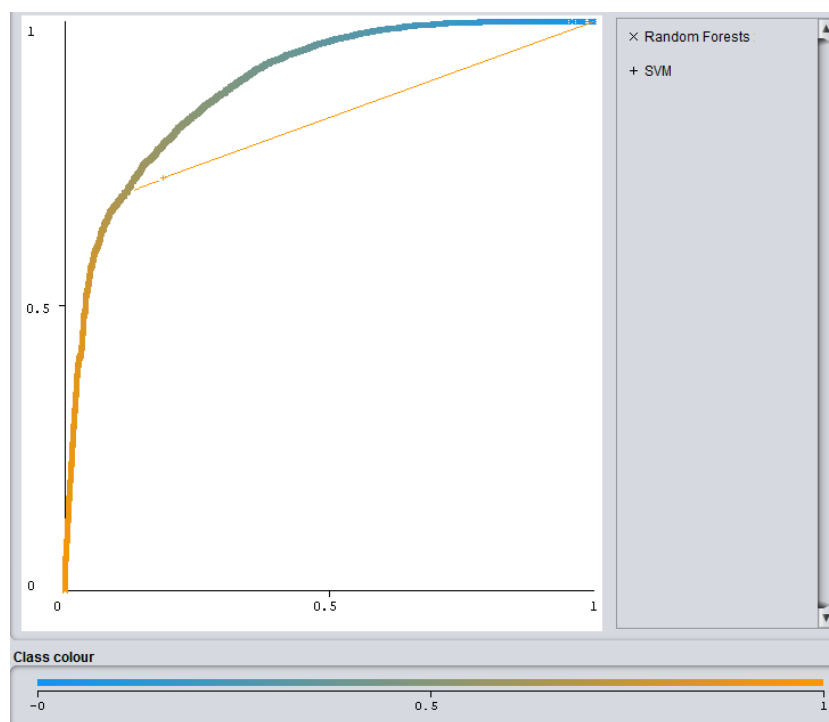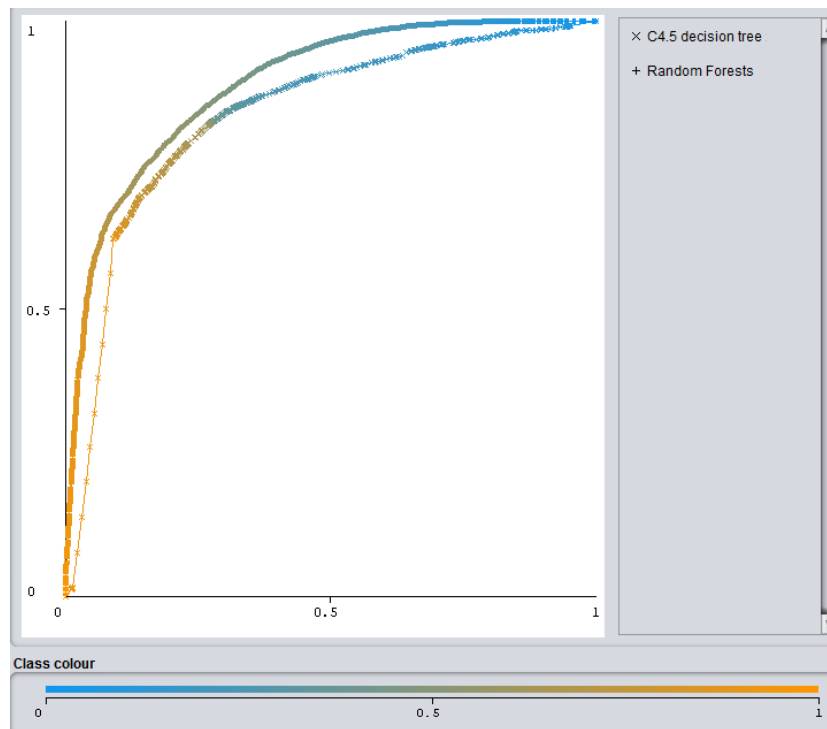**(A)** Random forests vs. Logistic ridge regression.

(B) Random forests vs. SVM.

(C) Random forests vs. C4.5 decision tree.

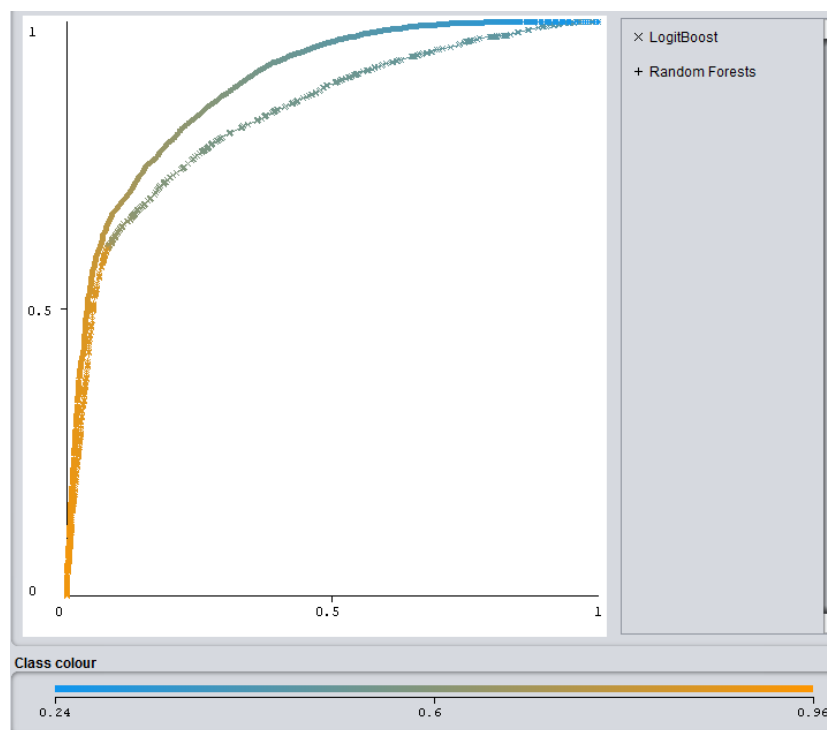**(D)** Random forests vs. LogitBoost.

Figure S3. Plots of receiver operating characteristic (ROC) curves using synthetic minority oversampling. (A) ROC curves representing random forests (the upper curve) and logistic ridge regression (the lower curve); (B) ROC curves representing random forests (the upper curve) and support vector machine (SVM) (the lower curve); (C) ROC curves representing random forests (the upper curve) and C4.5 decision tree (the lower curve); (D) ROC curves representing random forests (the upper curve) and LogitBoost (the lower curve). The vertical axis represents the true positive rate, and the horizontal axis represents the false positive rate. The "class colour" indicator near the bottom represents the threshold value for each instance. While the darker brown color indicates the higher threshold value, the darker blue color indicates the lower threshold value.