Supplementary Data

1.1 R programs

Data were analyzed using RStudio (http://www.rstudio.com/, Version 1.4.1106) with R statistical package (http://www.r-project.org/, Version 4.1.0).

In the least absolute shrinkage and selection operator (LASSO) regression, we used the R package "glmnet (https://cran.r-project.org/web/packages/glmnet/glmnet.pdf)" and tuned parameter λ after increasing by 0.001 between 0 and 1 using 10-fold cross-validation with grid search using "caret (https://cran.r-project.org/web/packages/caret/caret.pdf)." We used the R package "xgboost (https://cran.r-project.org/web/packages/xgboost/xgboost.pdf)" to perform eXtreme Gradient Boosting. In eXtreme Gradient Boosting, we tuned "nrounds" using 10-fold cross-validation and stopped the training with no improvement in the AUROC in test data after 5 rounds in a row. The parameter of "max_depth" was tuned between 1 and 10 with grid search, and the other parameters were fixed as follows: "eta," 0.1; "gamma," 0; "colsample_bytree," 1; "min_child_weight," 1; and "subsample," 1.

For internal validation, we used the R package "rsample (https://cran.r-project.org/web/packages/rsample/rsample.pdf)" using stratified 5-fold cross-validation.

The importance of variables was evaluated using the absolute value of standardized regression coefficient in the LASSO model and using Gain in the gradient-boosting decision tree model. In addition, to use the interpretable method, Shapley additive explanation (SHAP), the R package "xgboost" was applied.