

R programs

【Machine learning model】

Data were analyzed using RStudio (<http://www.rstudio.com/>, version 2022.02.3) with the R statistical language (<http://www.r-project.org/>, version 4.2.0). 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 five rounds in a row. The parameter “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.

【Prediction metrics】

The “pROC (<https://cran.r-project.org/web/packages/pROC/pROC.pdf>)” package was used to calculate sensitivity specificity using the youden-index method, and AUROC. The “PRROC (<https://cran.r-project.org/web/packages/PRROC/PRROC.pdf>)” package was used to calculate the AUPRC. The R package “CalibrationCurves (<https://github.com/BavoDC/CalibrationCurves>)” was used to calculate calibration slope and calibration intercept. As an internal validation, we used the R package “rsample (<https://cran.r-project.org/web/packages/rsample/rsample.pdf>)” to use stratified 5-fold cross-validation.

【SHAP method】

The R package “shapviz (<https://cran.r-project.org/web/packages/shapviz/shapviz.pdf>)” was used to calculate and visualize SHAP values.