给王老师做其他结题使用

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Supplementary Table 1. Diagnosis accuracy, sensitivity and specificity based on different classification methods with five-fold cross-validation | | | | | | |
|  | test | | | train | | |
|  | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | Accuracy |
| Logistic Regression | 0.77225 | 0.918535 | 0.844512 | 0.869441 | 0.935915 | 0.901339 |
| KNN | 0.540893 | 0.686644 | 0.614157 | 0.426959 | 0.617287 | 0.515706 |
| SVM | 0.850607 | 0.937032 | 0.893468 | 0.899937 | 0.919904 | 0.909881 |
| Neural Net | 0.821327 | 0.951418 | 0.885988 | 0.816869 | 0.900151 | 0.857395 |
| Random Forest | 0.89739 | 0.837338 | 0.867485 | 0.929279 | 0.913385 | 0.920938 |
| Bayes Tree | 0.837591 | 0.909593 | 0.873258 | 0.913360 | 0.946893 | 0.929994 |

The classification performances were based on age, gender, smoking status and all the 14 methylation sites we have detected. Sensitivity, specificity and accuracy were its mean value in 5-fold validations with 1,000 times repeats. SVM represent support vector machines and Kernel Methods. In the main body of the manuscript, sensitivity, specificity and accuracy were derived from training result of the classification.