**Supplementary Table 6. Model performances of NI and NIBG variables for predicting the presence of IPVD by different ML methods**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Dataset | AUROC | Sensitivity, % | Specificity, % | PPV, % | NPV, % | F1 score |
| NI model |
| AdaBoost | Training dataset | 0.85 (0.79, 0.91) | 88.9 (83.6, 94.2) | 63.0 (54.8, 71.1) | 78.3 (71.3, 85.2) | 79.1 (72.2,85.9) | 0.78 |
| Testing dataset | 0.81 (0.71, 0.91) | 88.9 (80.8, 97.0) | 63.6 (51.3, 76.0) | 80.0 (69.7, 90.3) | 77.8 (67.1,88.5) | 0.79 |
| GBDT | Training dataset | 0.88 (0.83, 0.94) a | 90.1 (85.1, 95.2) | 61.1 (52.9, 69.3) | 77.7 (70.6, 84.7) | 80.5 (73.8, 87.2) | 0.78 |
| Testing dataset | 0.83 (0.74, 0.93) | 86.1 (77.2, 95.0) | 54.5 (41.7, 67.4) | 75.6 (64.6, 86.7) | 70.6 (58.9, 82.3) | 0.73 |
| LR | Training dataset | 0.83 (0.77, 0.90) b | 71.6 (64.0, 79.2) | 81.5 (74.9, 88.0) | 85.3 (79.3, 91.3) | 65.7 (57.7, 73.7) | 0.76 |
| Testing dataset | 0.86 (0.77, 0.95) a | 77.8 (67.1, 88.5) | 77.3 (66.5, 88.1) | 84.8 (75.6, 94.1) | 68.0 (56.0, 80.0) | 0.78 |
| SVM | Training dataset | 0.88 (0.82, 0.93) c | 90.1 (85.1, 95.2) | 68.5 (60.7, 76.4) | 81.1 (74.5, 87.7) | 82.2 (75.8, 88.7) | 0.82 |
| Testing dataset | 0.84 (0.74, 0.93) | 91.7 (84.6, 98.8) | 63.6 (51.3, 76.0) | 80.5 (72.5, 92.2) | 82.4 (70.3, 90.7) | 0.80 |
| NIBG model |
| AdaBoost | Training dataset | 0.95 (0.92, 0.99） | 91.4 (86.6, 96.1） | 81.5 (74.9, 88.0） | 88.1 (82.6, 93.6) | 86.3 (80.5, 92.1） | 0.87 |
| Testing dataset | 0.88 (0.80, 0.96) | 86.1 (77.2, 95.0） | 81.8 (71.9, 91.7） | 88.6 (80.4, 96.8） | 78.3 (67.6, 88.9） | 0.85 |
| GBDT | Training dataset | 0.99 (0.97, 1.00) a | 97.5 (94.9, 100.0) | 88.9 (83.6, 94.2) | 92.9 (88.6, 97.3) | 96.0 (92.7, 99.3) | 0.94 |
| Testing dataset | 0.83 (0.74, 0.93) | 88.9 (80.8, 97.0) | 72.7 (61.3, 84.2) | 84.2 (74.8, 93.6) | 80.0 (69.7, 90.3) | 0.83 |
| LR | Training dataset | 0.87 (0.82, 0.93) ab | 75.3 (68.0, 82.6) | 83.3 (77.0, 89.6) | 87.1 (81.5, 92.8) | 69.2 (61.4, 77.0) | 0.79 |
| Testing dataset | 0.90 (0.82, 0.98) | 83.3 (73.7, 92.9) | 86.4 (77.5, 95.2) | 90.9 (83.5, 98.3) | 76.0 (65.0, 87.0) | 0.85 |
| SVM | Training dataset | 0.85 (0.79, 0.91) ab | 82.7 (76.3, 89.1） | 72.2 (64.7, 79.8） | 81.7 (75.2, 88.2） | 73.6 (66.1, 81.0) | 0.78 |
| Testing dataset | 0.82 (0.72, 0.92) a | 75.0 (63.9, 86.1） | 81.8 (71.9, 91.7） | 87.1 (78.5, 95.7） | 66.7 (54.5, 78.8） | 0.78 |

Statistical quantifications were demonstrated with 95% CI, when applicable· AUROCs of models based on each algorithm were statistically compared with that of the other, for the same strategy (a vs. AdaBoost, *p*<0.05; bvs*.* GBDT, *p*<0.05; c vs*.* LR, *p*<0.05). LR, logistic regression; SVM, support vector machine. Initially, we tried to apply four different ML methods to predict the cirrhotic patients’ presence of IPVD. After data preprocessing, feature selection and model construction were conducted on the training dataset by the four ML methods, respectively. Then model evaluation was constructed on the testing dataset. Our data showed that the AUCROC of the NI model by LR was 0.86 (95% CI: 0.74-0.93), which indicated a moderate ability of discrimination. Meanwhile, the sensitivity and specificity of the NI model by LR were lower than 80.0%, indicating a probable a relative low accuracy. Though the AUCROC of NIBG by LR was higher than 0.90 on the testing dataset, the AUCROC of NIBG on the training dataset was lower than that on the testing dataset, suggesting the probability of underfitting. Moreover, the ABG analysis results were not available for every patient. Our preliminary experiment showed that one-step model using noninvasive variables based on machine learning methods was unable to meet the requirements of preliminary screening from large amounts of patients.