|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The discrimination and calibration performance of different candidate strategy models for identifying the best-performing ML algorithm to develop the ML system** | | | | |
|  | **Predicting poor responders** | | **Predicting hyper-responders** | |
|  | **AUROC (95% CI)** | **Brier score** | **AUROC (95% CI)** | **Brier score** |
| XGBoost | 0.930 (0.918, 0.941) | 0.064 | 0.867 (0.851, 0.883) | 0.071 |
| Lasso logistic regression | 0.924 (0.911, 0.935) | 0.067 | 0.863 (0.846, 0.880) | 0.072 |
| Ridge logistic regression | 0.920 (0.908, 0.932) | 0.068 | 0.859 (0.842, 0.876) | 0.072 |
| SVM-RBF | 0.915 (0.902, 0.927) | 0.072 | 0.847 (0.830, 0.864) | 0.086 |
| Random Forest | 0.925 (0.912, 0.937) | 0.066 | 0.848 (0.831, 0.866) | 0.072 |
| MLP | 0.919 (0.905, 0.932) | 0.121 | 0.861 (0.844, 0.877) | 0.121 |
| XGBoost with only top features | 0.929 (0.917, 0.940) | 0.065 | 0.869 (0.853, 0.884) | 0.07 |
| AUROC, area under the receiver operating characteristic curve; CI, confidence interval.  CIs are the 95% ranges among 2000 bootstrapped replicates. | | | | |