**Table 3 |** Predictive performance of prediction participants with any psychiatric diagnosis at follow-up in training and modeling details for tested feature sets

| Feature Set | PR-AUC (cross-validated) | 95% CI | Significant (vs. baseline) | Feature Selection Method | Model Description | PR AUC (held-out test set) |
|---|---|---|---|---|---|---|
| Top 20 A1 - Biology + psychometry | 0.801 | [0.733,0.866] | Yes | None | SVM (C-SVC, Polynomial Kernel; cost = 0.001, $\gamma$ = 0.01, degree = 2) | 0.700 |
| Top 15 A1 - Biology | 0.780 | [0.714,0.845] | Yes | None | Ridge Logistic Regression ($\lambda$ = 100.0) | NA |
| Top 5 A1 - Psychometry | 0.782 | [0.710,0.848] | Yes | None | SVM (C-SVC, Polynomial Kernel; cost = 0.001, $\gamma$ = 1.0, degree = 2) | NA |
| Top 40 A1+A5 - Biology + psychometry | 0.782 | [0.707,0.850] | Yes | None | SVM (C-SVC, Gaussian Kernel; cost = 0.01, $\gamma$ = 0.1) | NA |
| Top 30 A1+A5 - Biology | 0.772 | [0.706,0.833] | Yes | None | Random Forest (1000 trees, deviance criterion, min leaf size = 5, splits = 1, $\alpha$ = 1, variables to split = 1.291√n) | NA |
| Top 10 A1+A5 - Psychometry | 0.753 | [0.680,0.822] | Yes | None | SVM (C-SVC, Gaussian Kernel; cost = 0.001, $\gamma$ = 0.01) | NA |
| All features - Biology + psychometry | 0.714 | [0.651,0.772] | Yes | SES (maxK = 2, $\alpha$ = 0.1 and budget = 3 * nvars) | SVM (C-SVC, Polynomial Kernel; cost = 0.001, $\gamma$ = 1.0, degree = 2) | 0.583 |
| All features - Biology | 0.728 | [0.658,0.792] | Yes | LASSO (penalty = 0.0) | SVM (C-SVC, Polynomial Kernel; cost = 0.01, $\gamma$ = 0.01, degree = 3) | NA |
| All features - Psychometry | 0.737 | [0.657,0.807] | Yes | LASSO (penalty = 0.25) | SVM (C-SVC, Gaussian Kernel; cost = 1.0, $\gamma$ = 0.001) | NA |