COVID-19 Optimal Surrogate Analysis Report (Moderna mock study)

Table 1: All learner-screen combinations (14 in total) used as input to the Superlearner.

Learner	Screen
SL.mean	all
SL.glmnet.0	all
SL.glmnet.1	all
SL.xgboost.2.no	all
SL.xgboost.4.no	all
SL.xgboost.2.yes	all
SL.xgboost.4.yes	all
SL.ranger.yes	all
SL.ranger.no	all
SL.glm	all
SL.glm	glmnet
SL.glm	univar_logistic_pval
SL.glm	highcor_random

Table 2: All variable sets (13 in total) with immunological markers for which Superlearner was run.

Variable Set Name	**Variables included in the set**
1_noisyVars	Noisy variables only (3 random predictors based off gaussian distribution)
2_baselineRiskFactors	Baseline risk factors only (Reference model)
3_varset_bAbSpike	Baseline risk factors + bAb anti-Spike markers
4_varset_bAbRBD	Baseline risk factors + bAb anti-RBD markers
5_varset_pnabID50	Baseline risk factors + p-nAb ID50 markers
6_varset_pnabID80	Baseline risk factors + p-nAb ID80 markers
7_varset_lnabMN50	Baseline risk factors + l-nAb MN50 markers
8_varset_bAb_pnabID50	Baseline risk factors + bAb markers + p-nAb ID50 markers
9_varset_bAb_pnabID80	Baseline risk factors + bAb markers + p-nAb ID80 markers
10_varset_bAb_lnabMN50	Baseline risk factors + bAb markers + l-nAb MN50 markers
11_varset_bAb_combScores	Baseline risk factors + bAb markers + combination scores across the five markers [PCA1, PCA2, FSDAM1/FSDAM2 (the first two components of nonlinear PCA), and the maximum signal diversity score]
12_varset_allMarkers	Baseline risk factors + all individual markers
13_varset_allMarkers_combScores	Baseline risk factors + all individual markers + all combination scores (Full model)

Note:
Baseline risk factors include risk score, HighRiskInd and MinorityInd variables.

Table 3: Superlearner performance across all 13 variable sets sorted by weighted CV-AUC performance.

Variable set	CV-AUC (95% CI)
11_varset_bAb_combScores	0.724 [0.631, 0.817]
10_varset_bAb_lnabMN50	0.703 [0.615, 0.792]
9_varset_bAb_pnabID80	0.696 [0.606, 0.787]
13_varset_allMarkers_combScores	0.695 [0.604, 0.787]
3_varset_bAbSpike	0.690 [0.601, 0.780]
8_varset_bAb_pnabID50	0.690 [0.602, 0.778]
7_varset_lnabMN50	0.683 [0.593, 0.773]
12_varset_allMarkers	0.682 [0.594, 0.770]
2_baselineRiskFactors	0.678 [0.591, 0.765]
4_varset_bAbRBD	0.648 [0.551, 0.746]
5_varset_pnabID50	0.646 [0.556, 0.735]
6_varset_pnabID80	0.637 [0.544, 0.730]
1_noisyVariables	0.522 [0.421, 0.624]

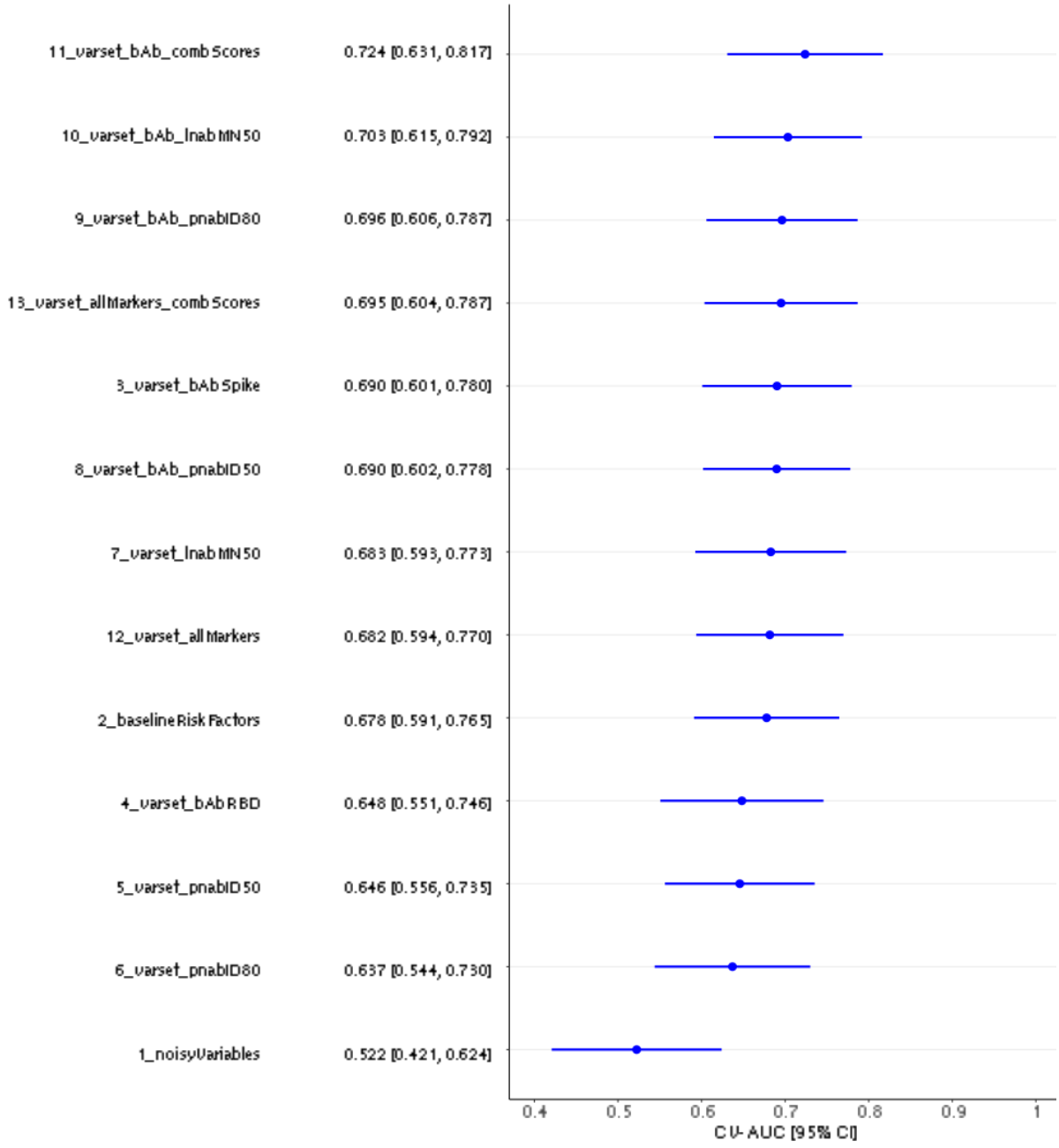

Figure 1: Forest plot showing Superlearner performance (weighted CV-AUC with 95% CI) across all 13 variable sets.

Appendix

Forest plots, ROC curves and predicted probability plots are shown for each variable set and arranged according to Superlearner performance.

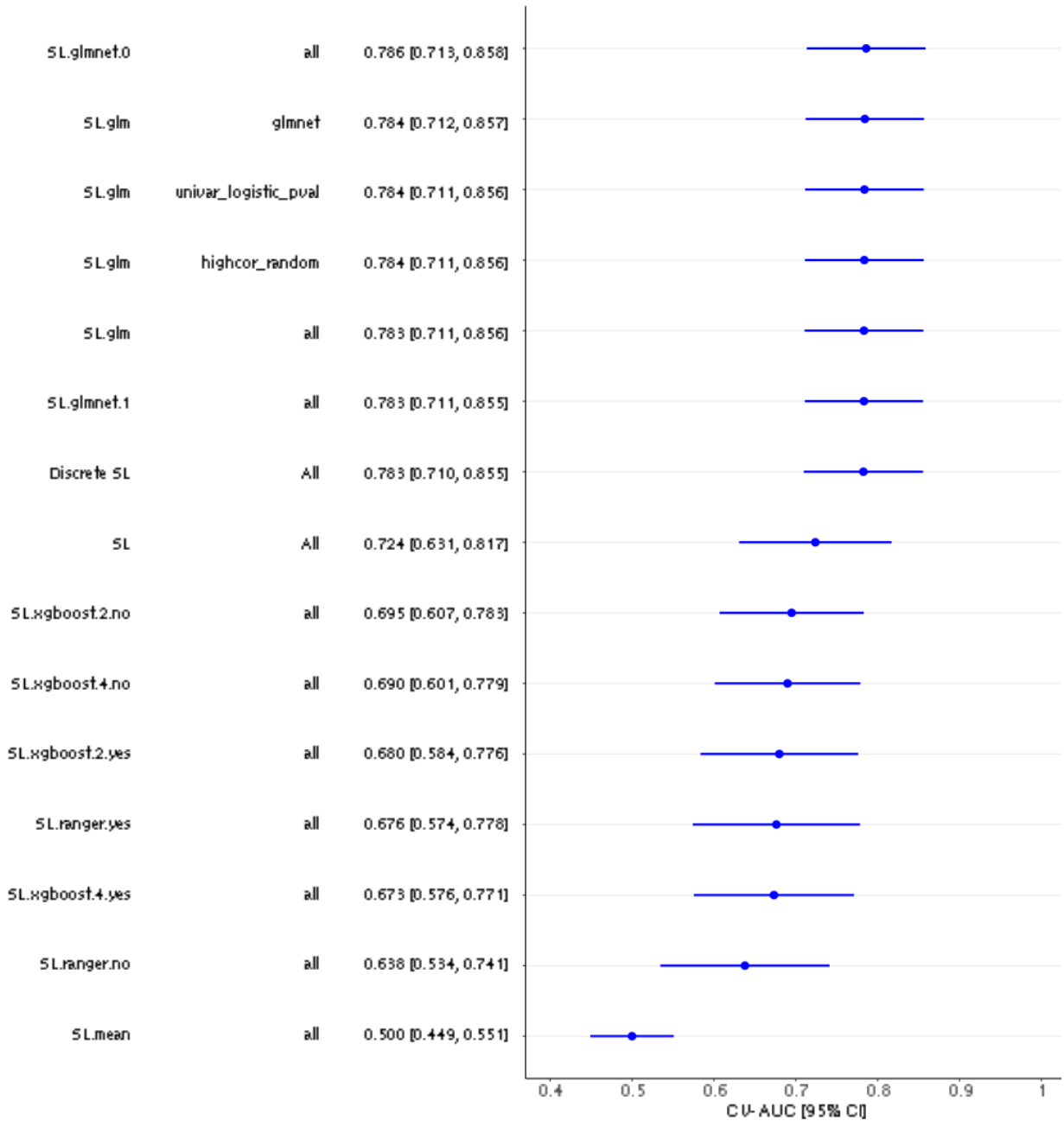

Figure 2: 11_varset_bAb_combScores: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

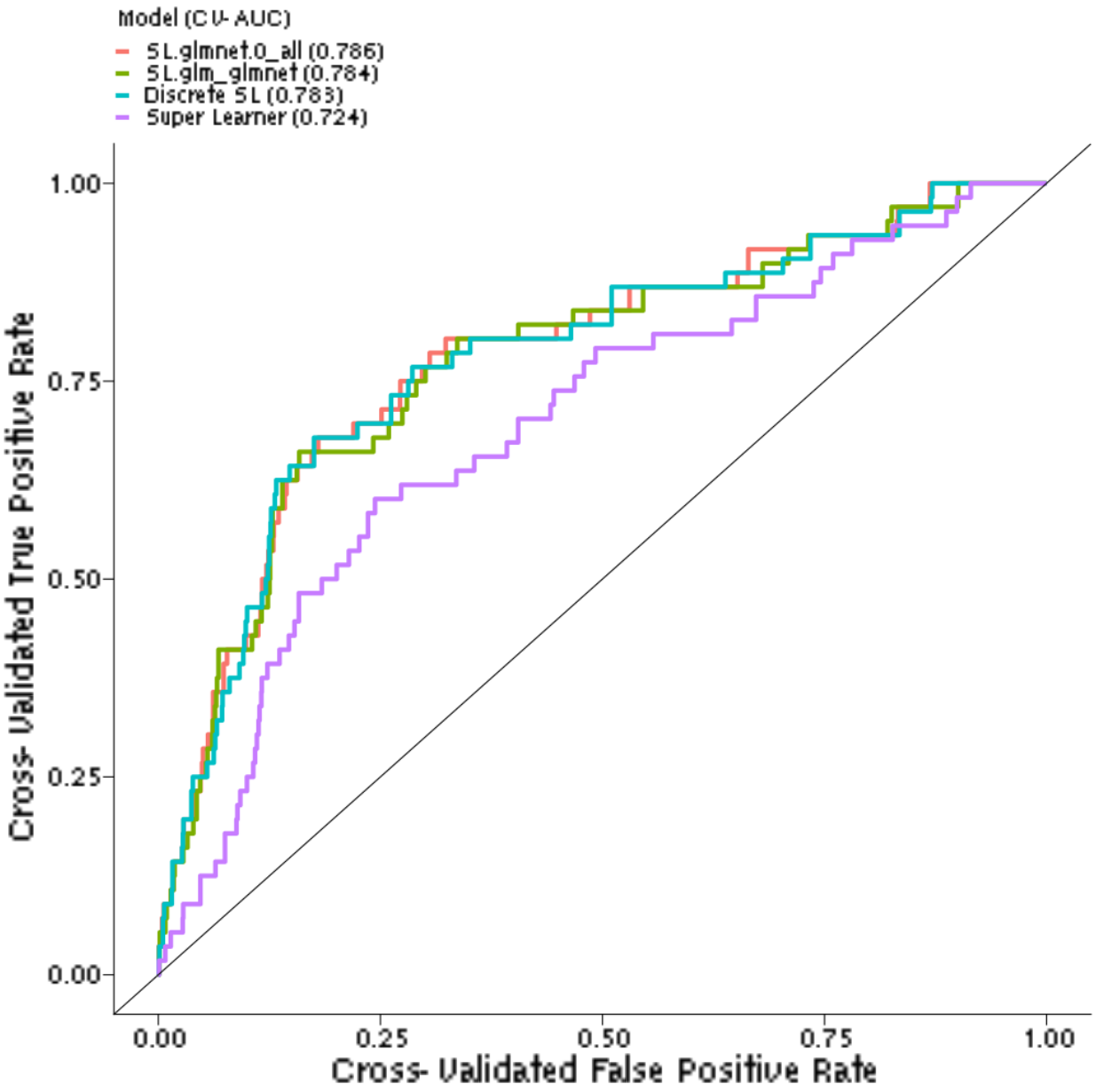

Figure 3: 11_varset_bAb_combScores: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

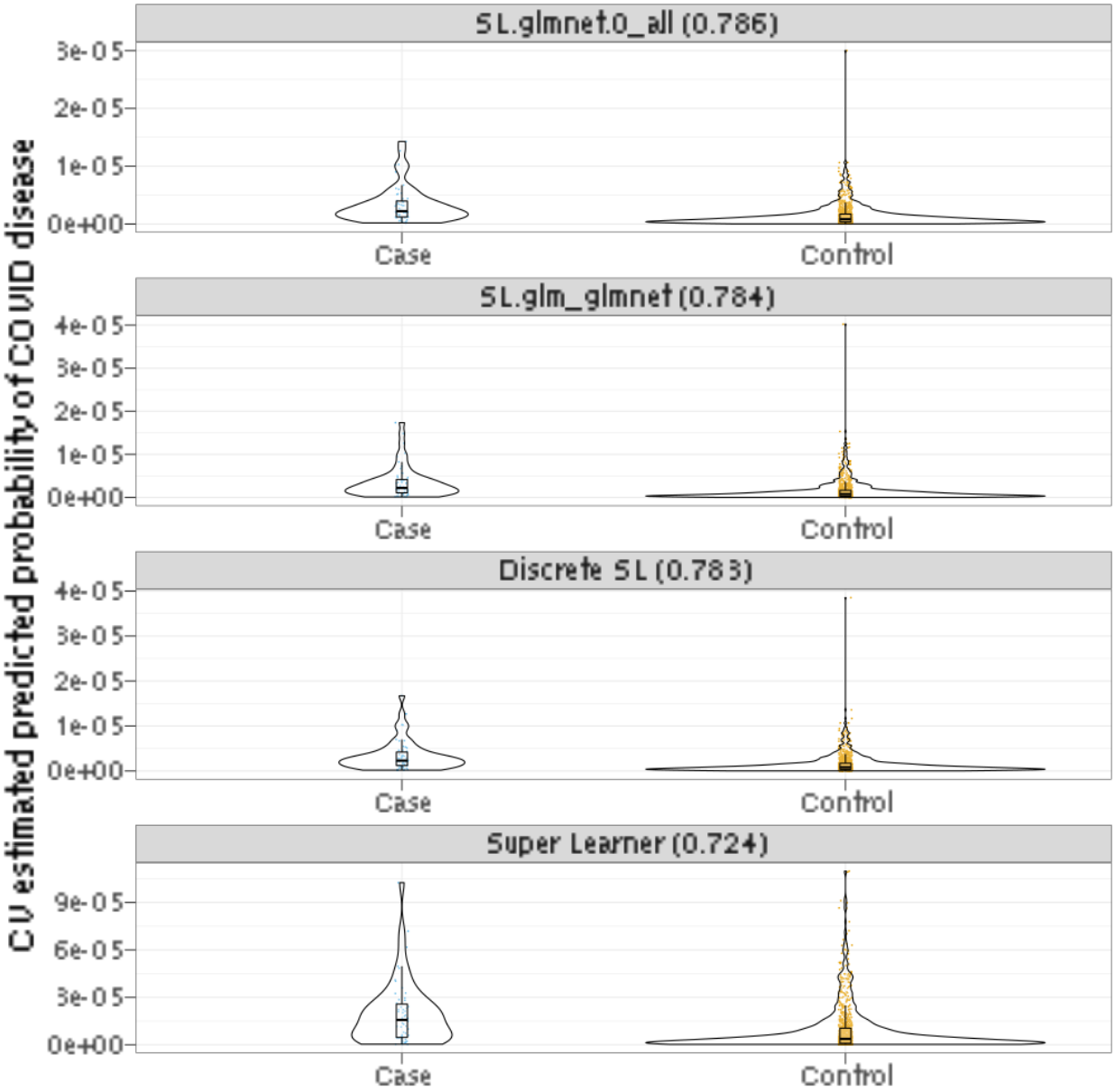

Figure 4: 11_varset_bAb_combScores: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

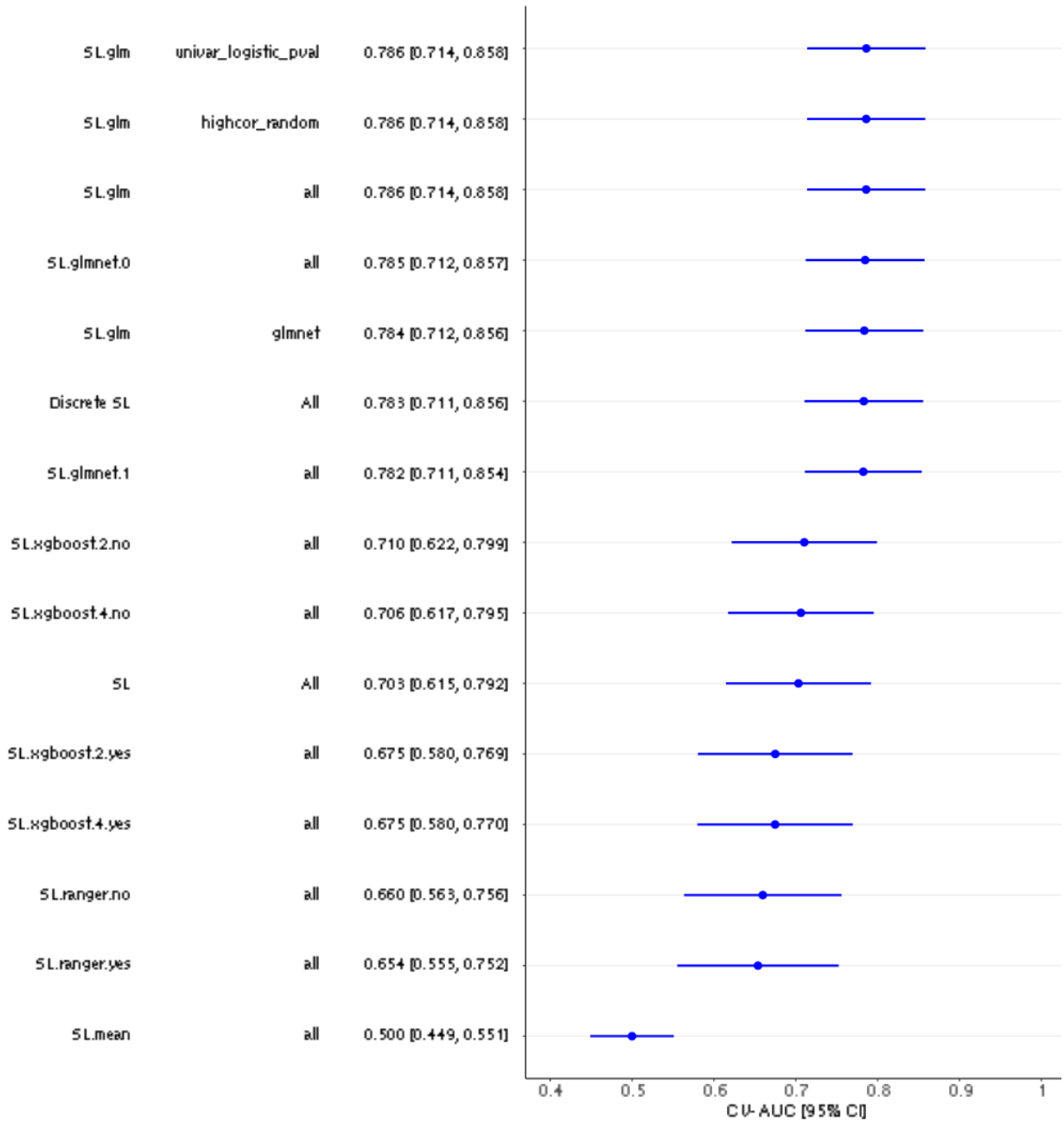

Figure 5: 10_varset_bAb_lnabMN50: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

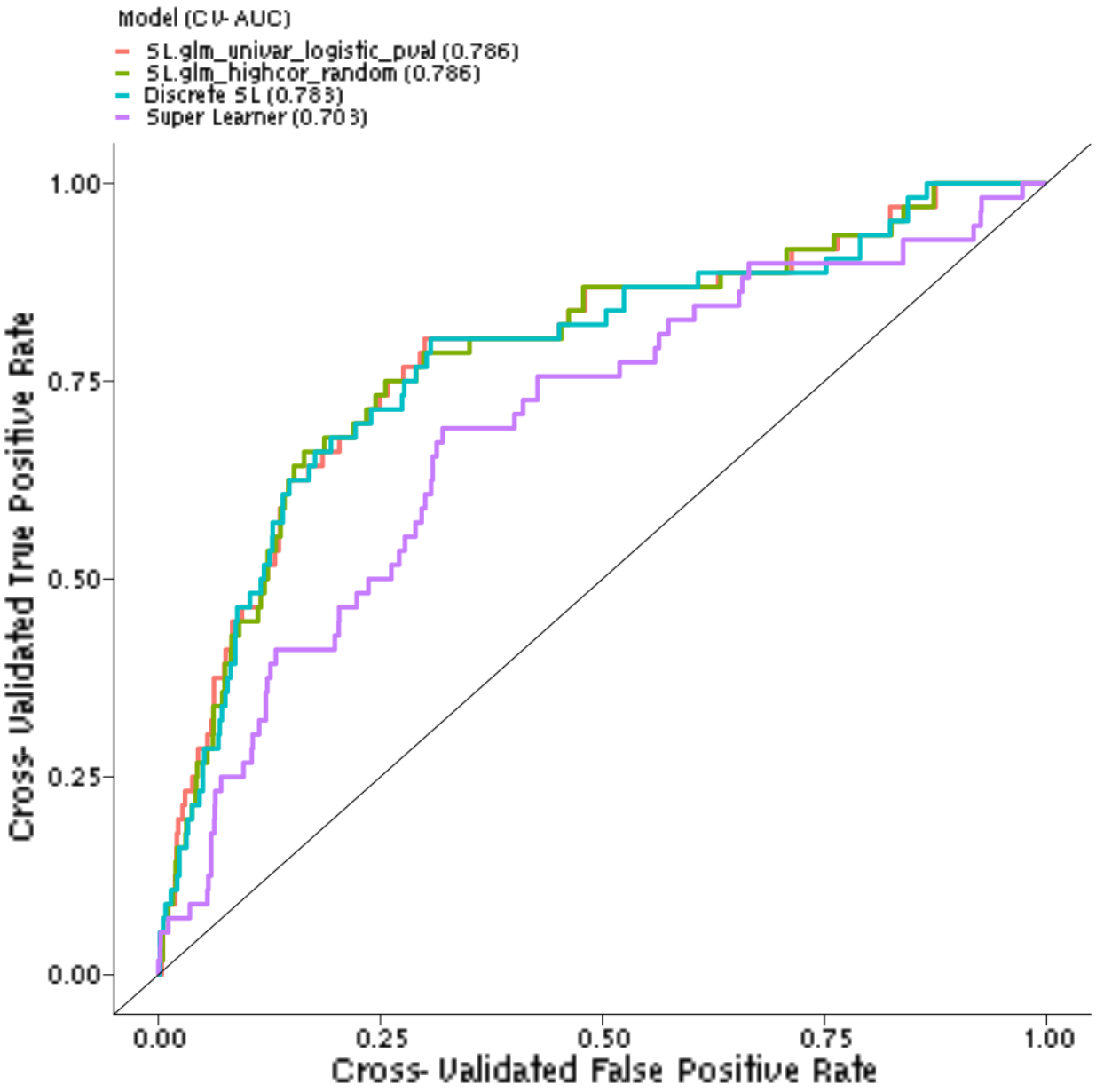

Figure 6: 10_varset_bAb_lnabMN50: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

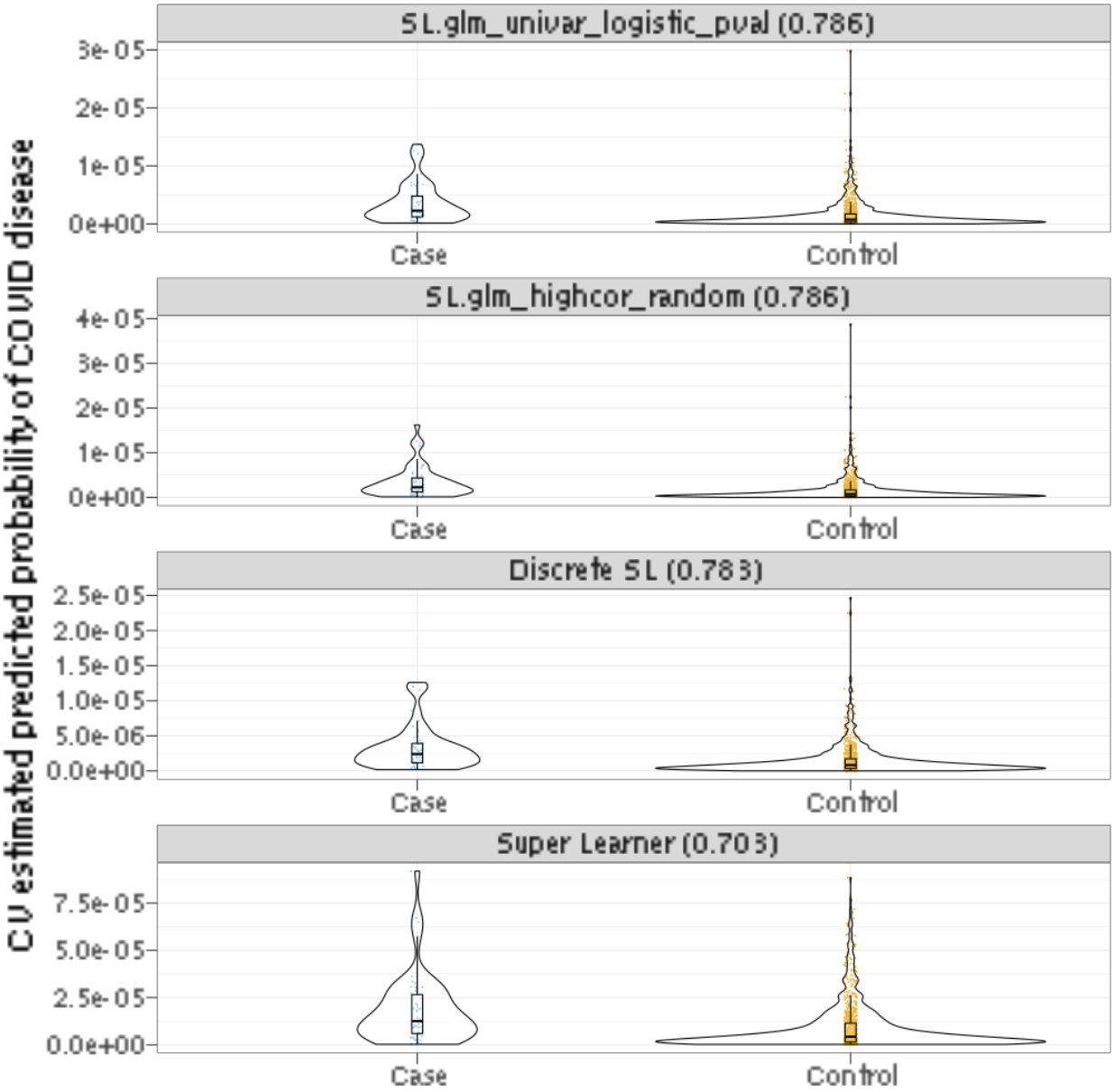

Figure 7: 10_varset_bAb_lnabMN50: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

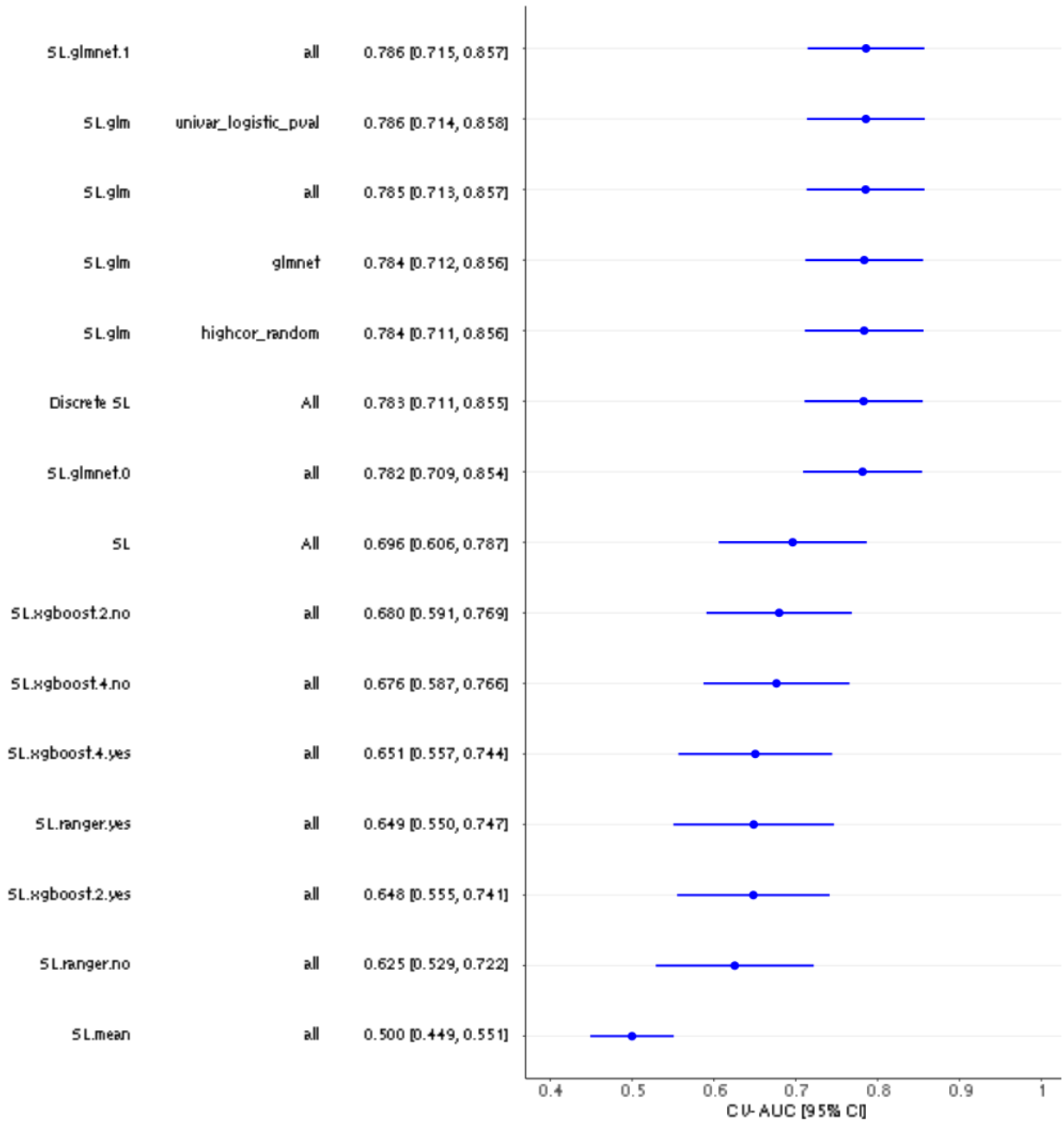

Figure 8: 9_varset_bAb_pnabID80: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

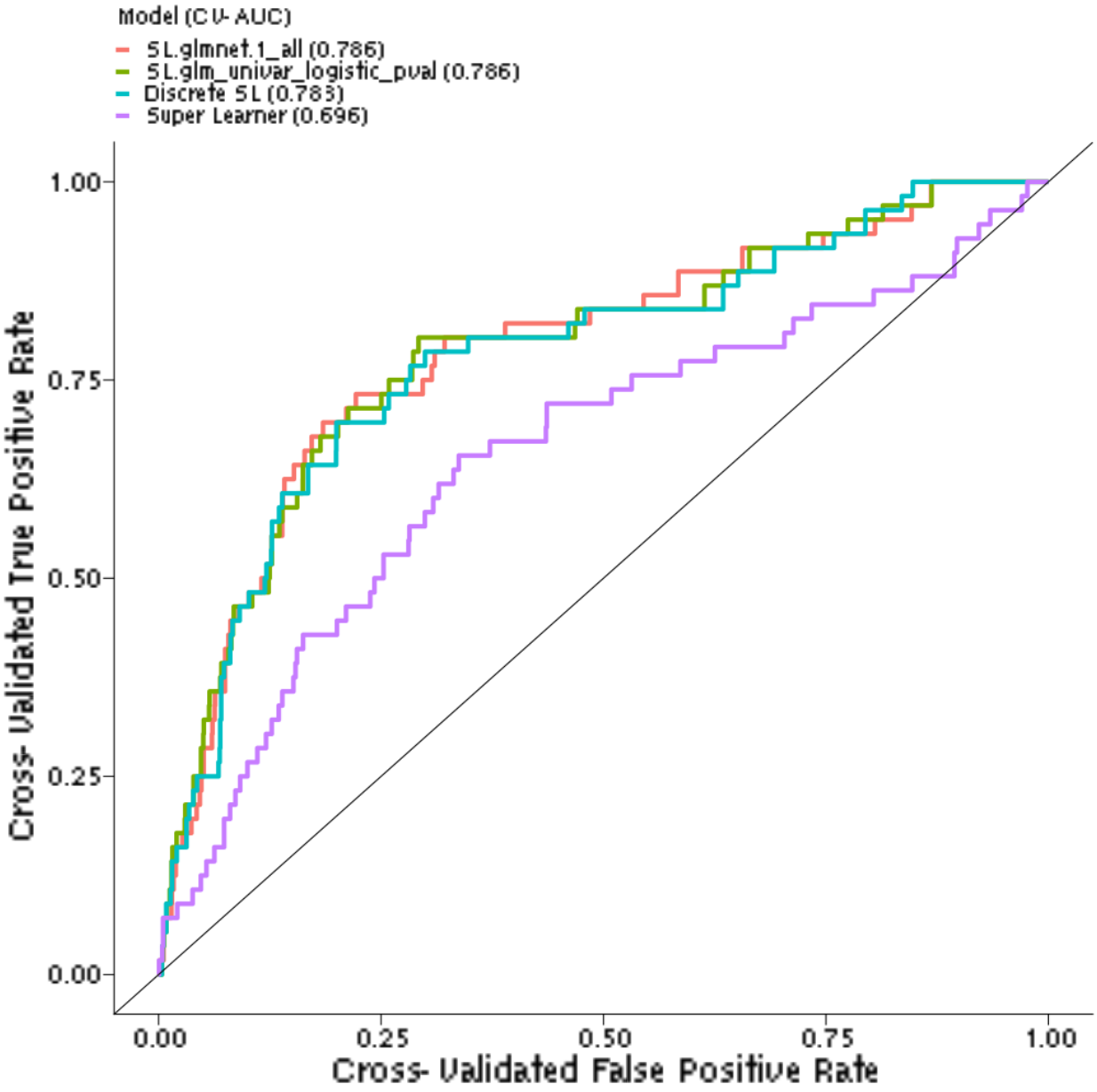

Figure 9: 9_varset_bAb_pnabID80: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

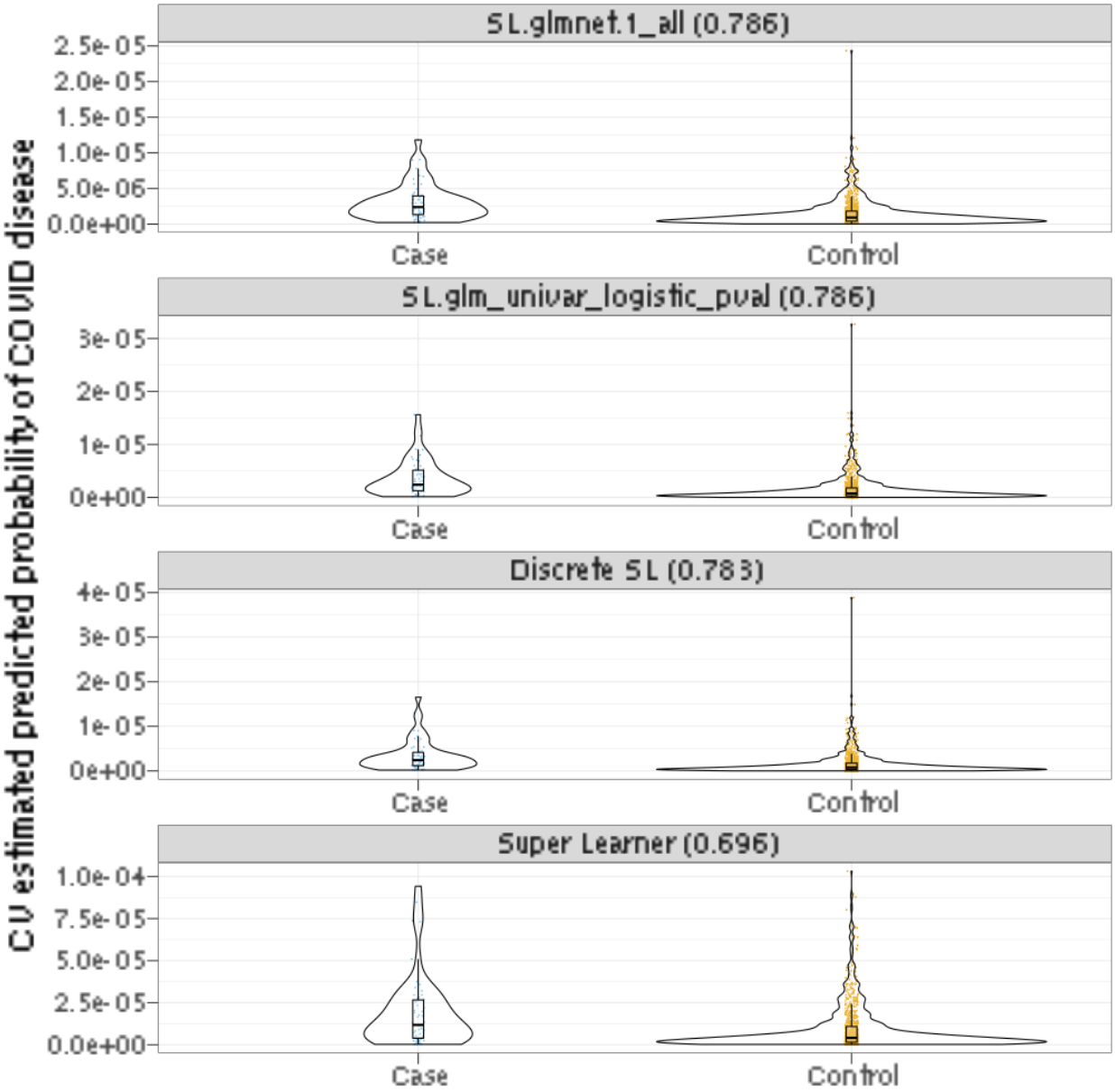

Figure 10: 9_varset_bAb_pnabID80: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

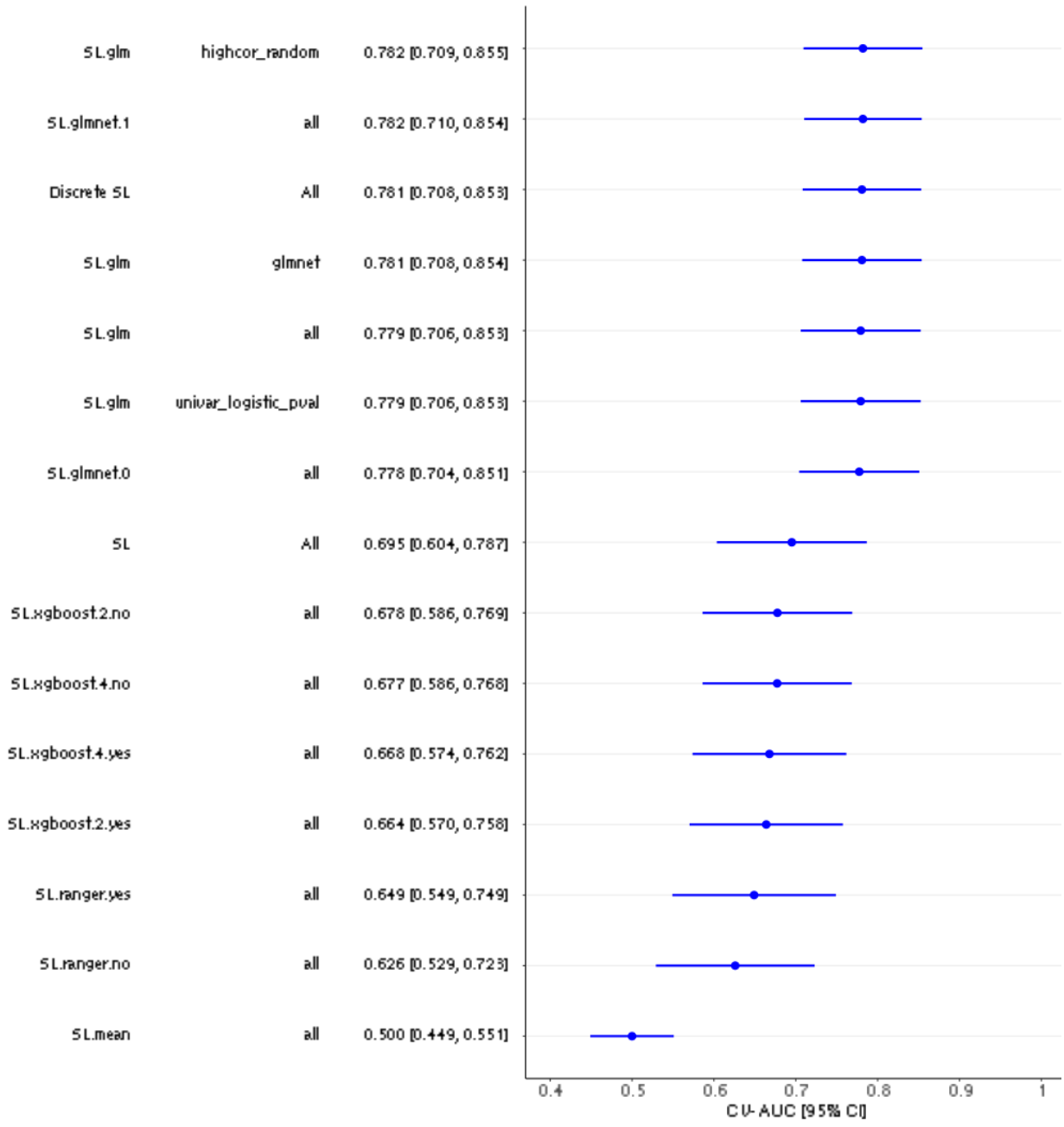

Figure 11: 13_varset_allMarkers_combScores: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

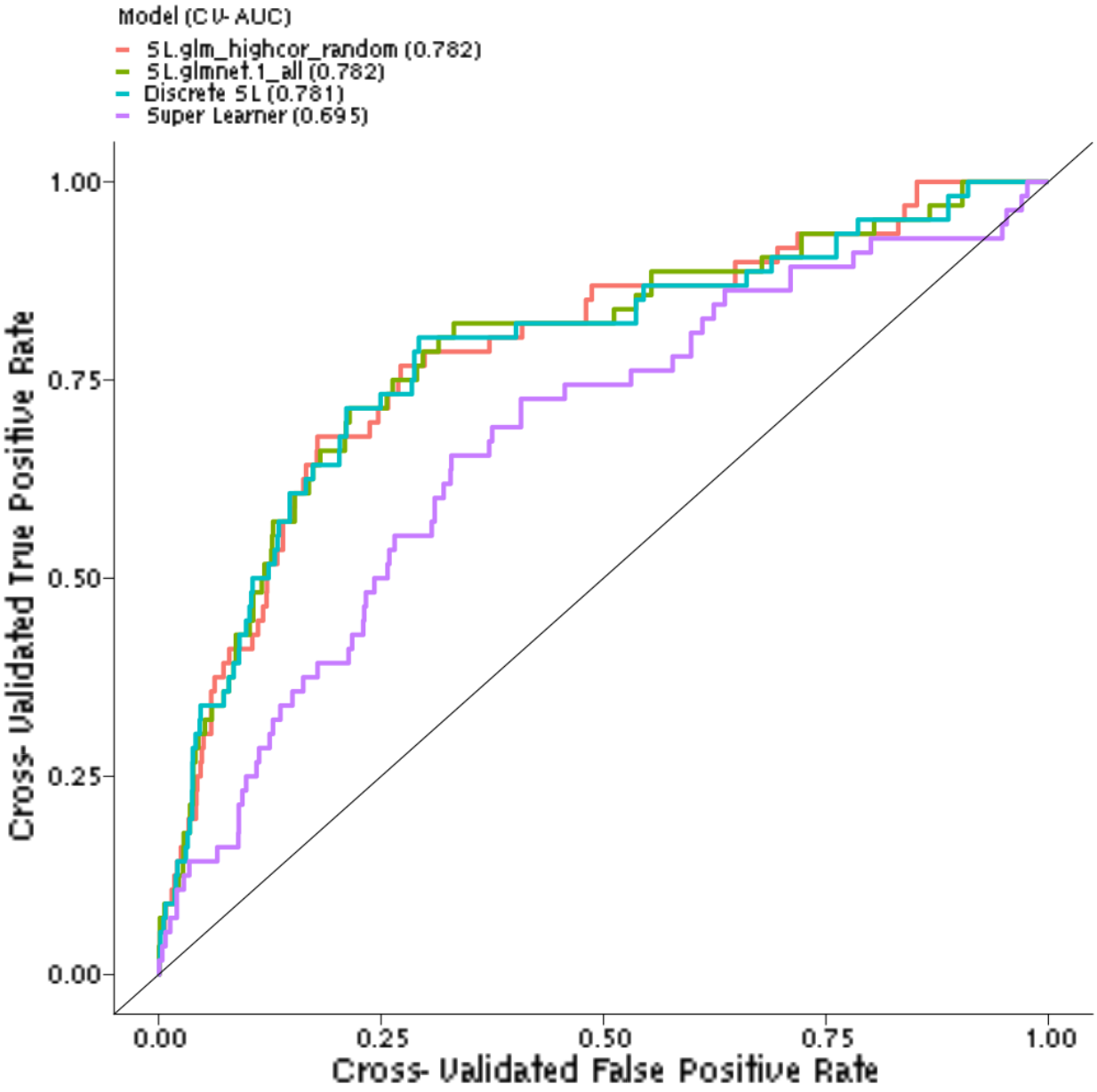

Figure 12: 13_varset_allMarkers_combScores: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

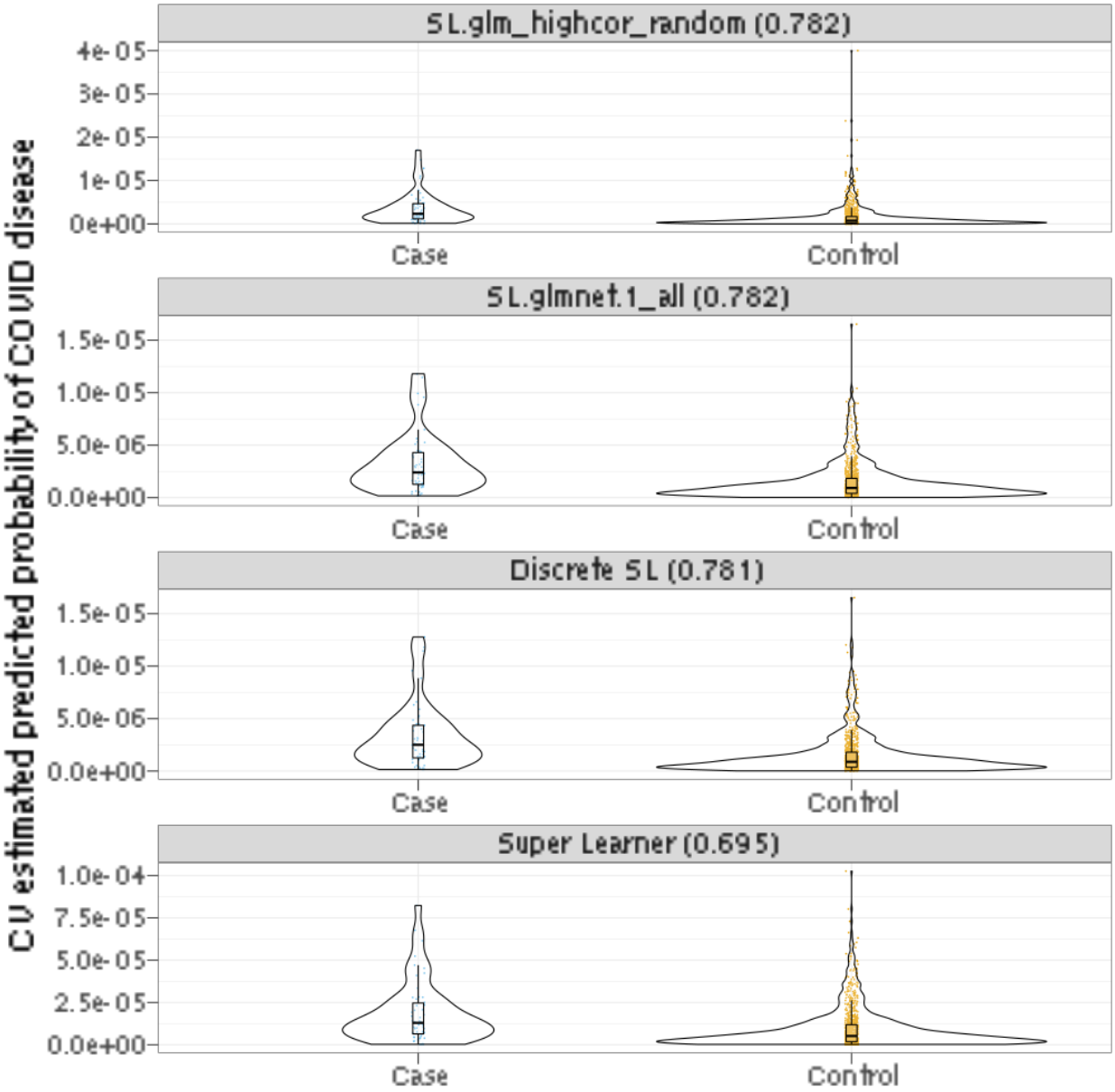

Figure 13: 13_varset_allMarkers_combScores: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

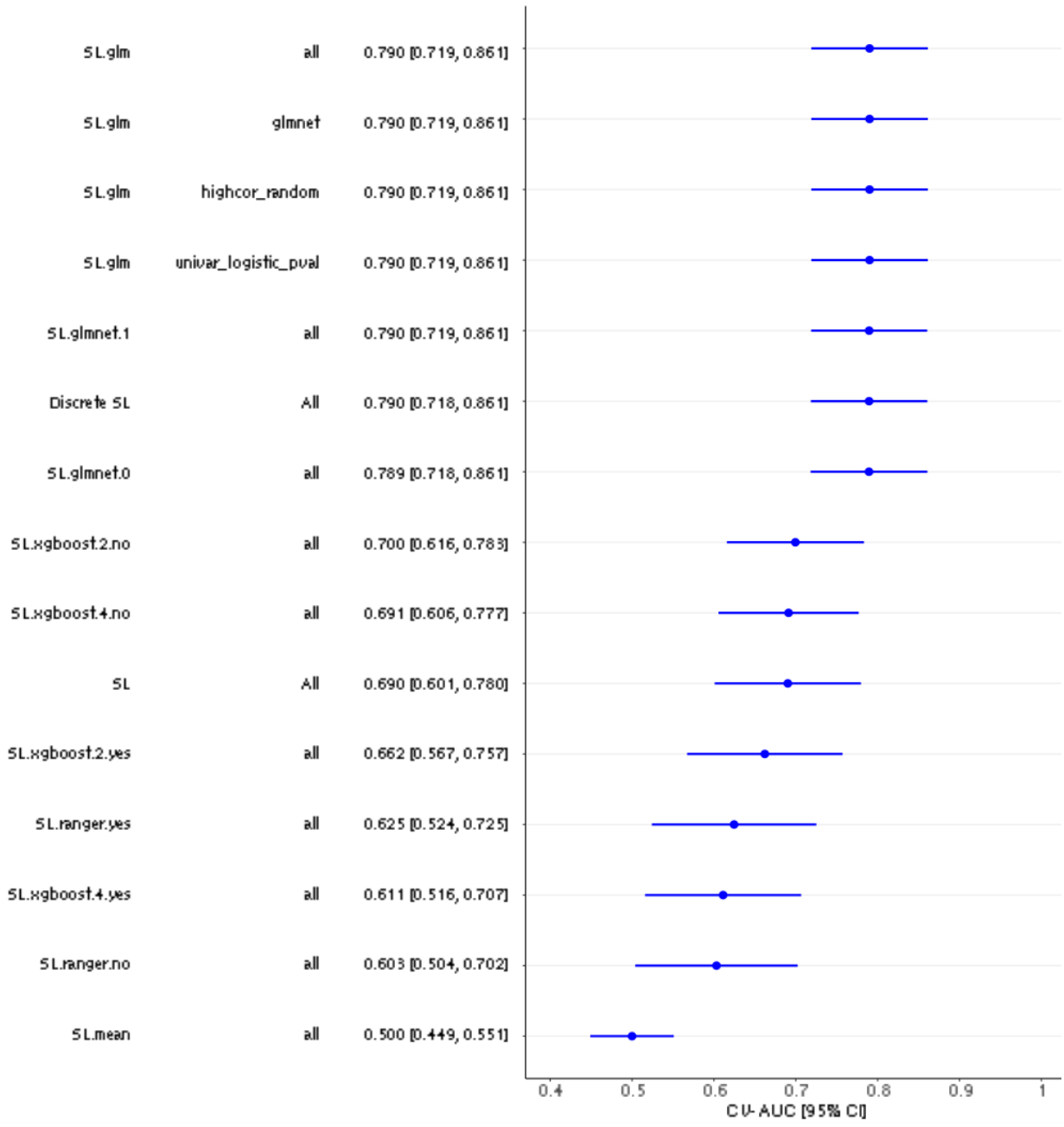

Figure 14: 3_varset_bAbSpike: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

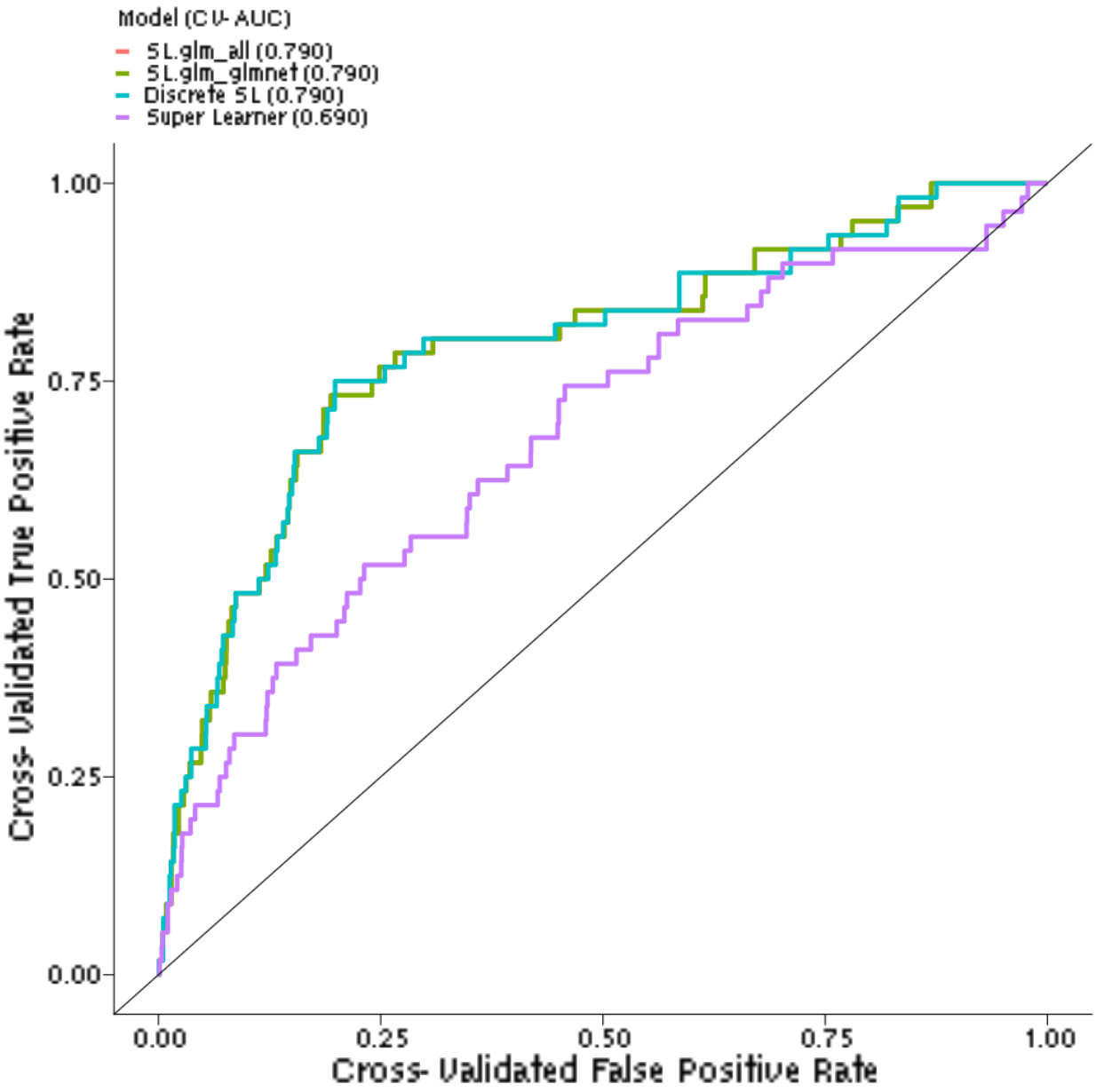

Figure 15: 3_varset_bAbSpike: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

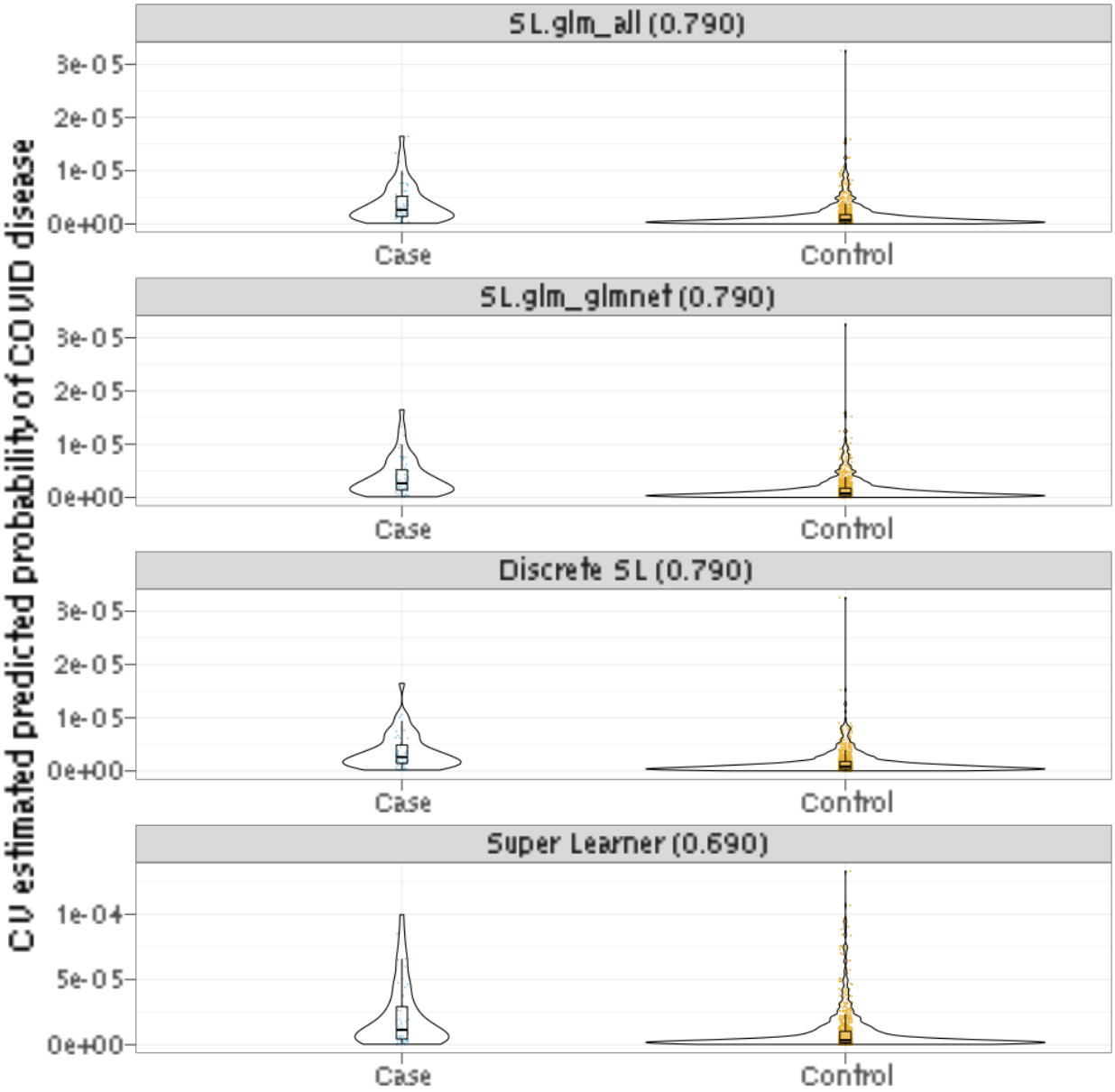

Figure 16: 3_varset_bAbSpike: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

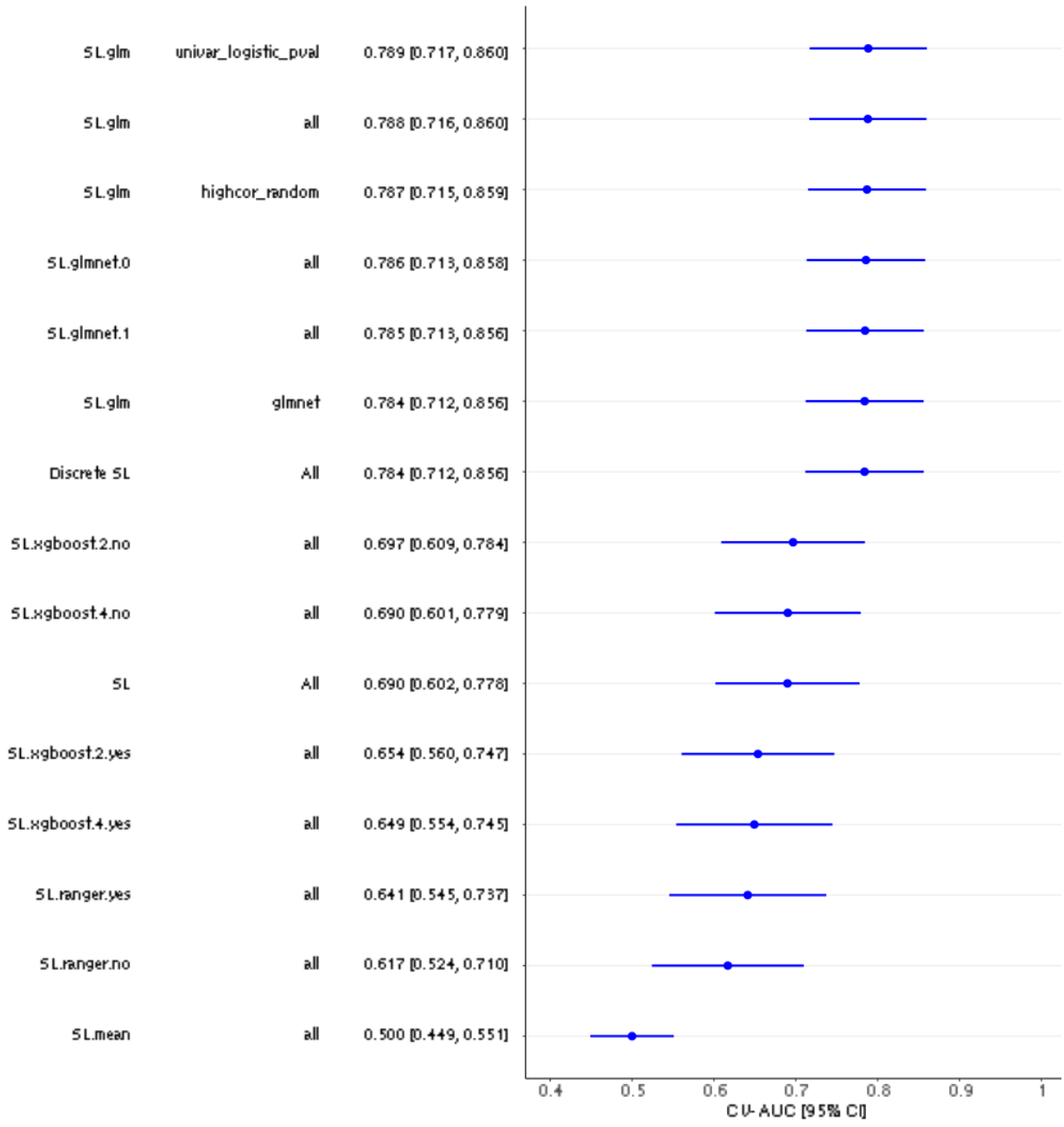

Figure 17: 8_varset_bAb_pnabID50: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

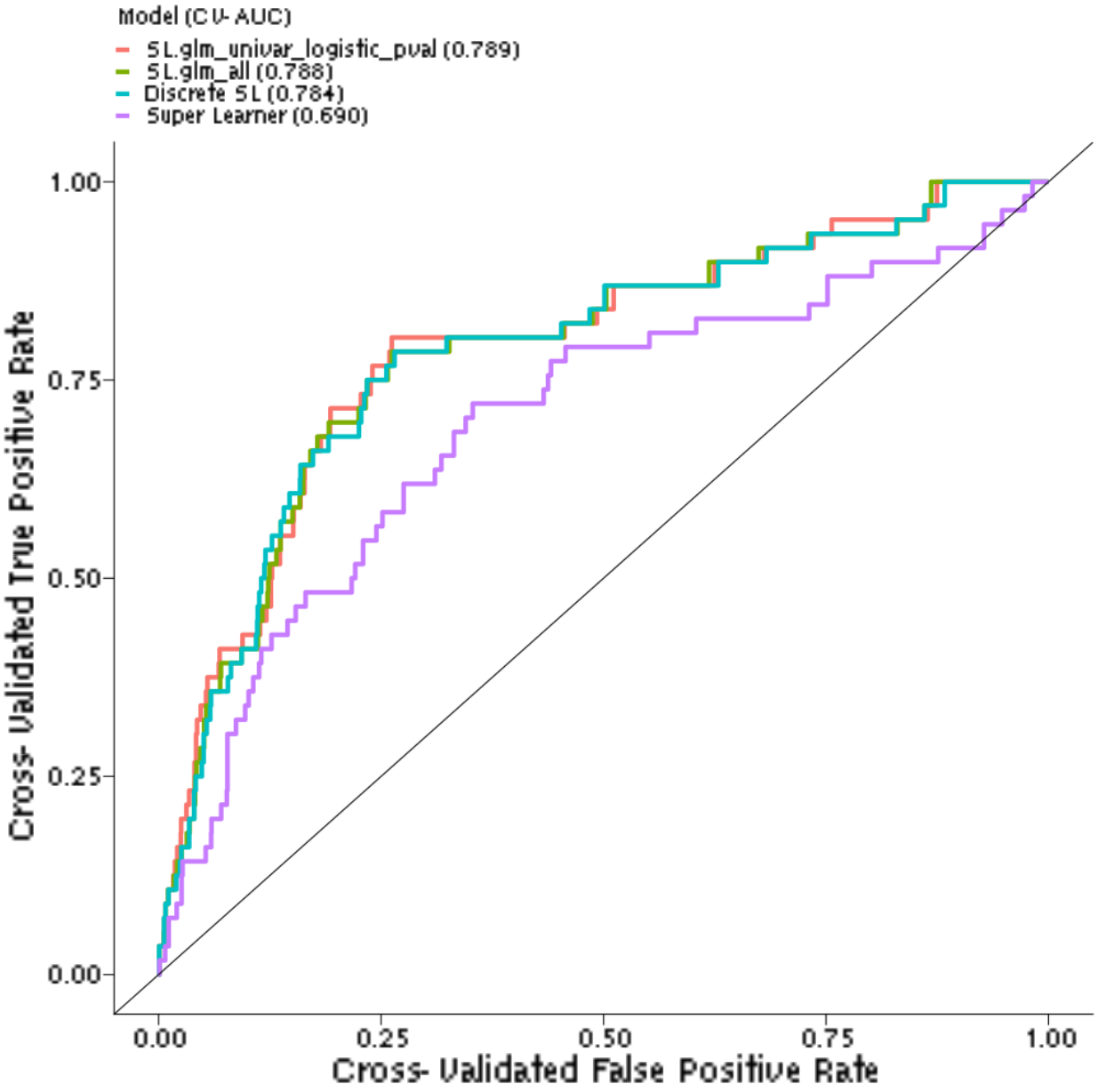

Figure 18: 8_varset_bAb_pnabID50: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

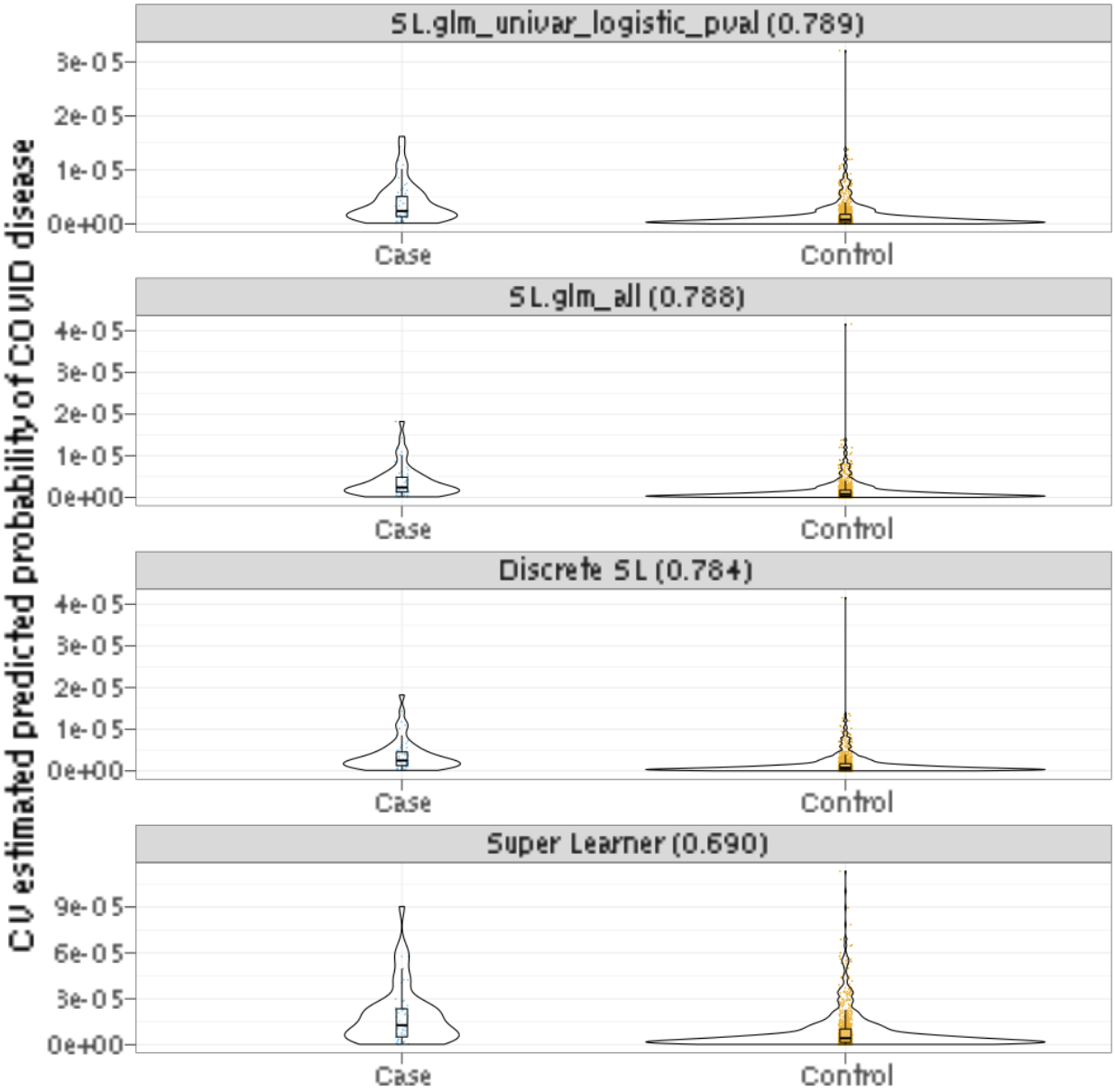

Figure 19: 8_varset_bAb_pnabID50: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

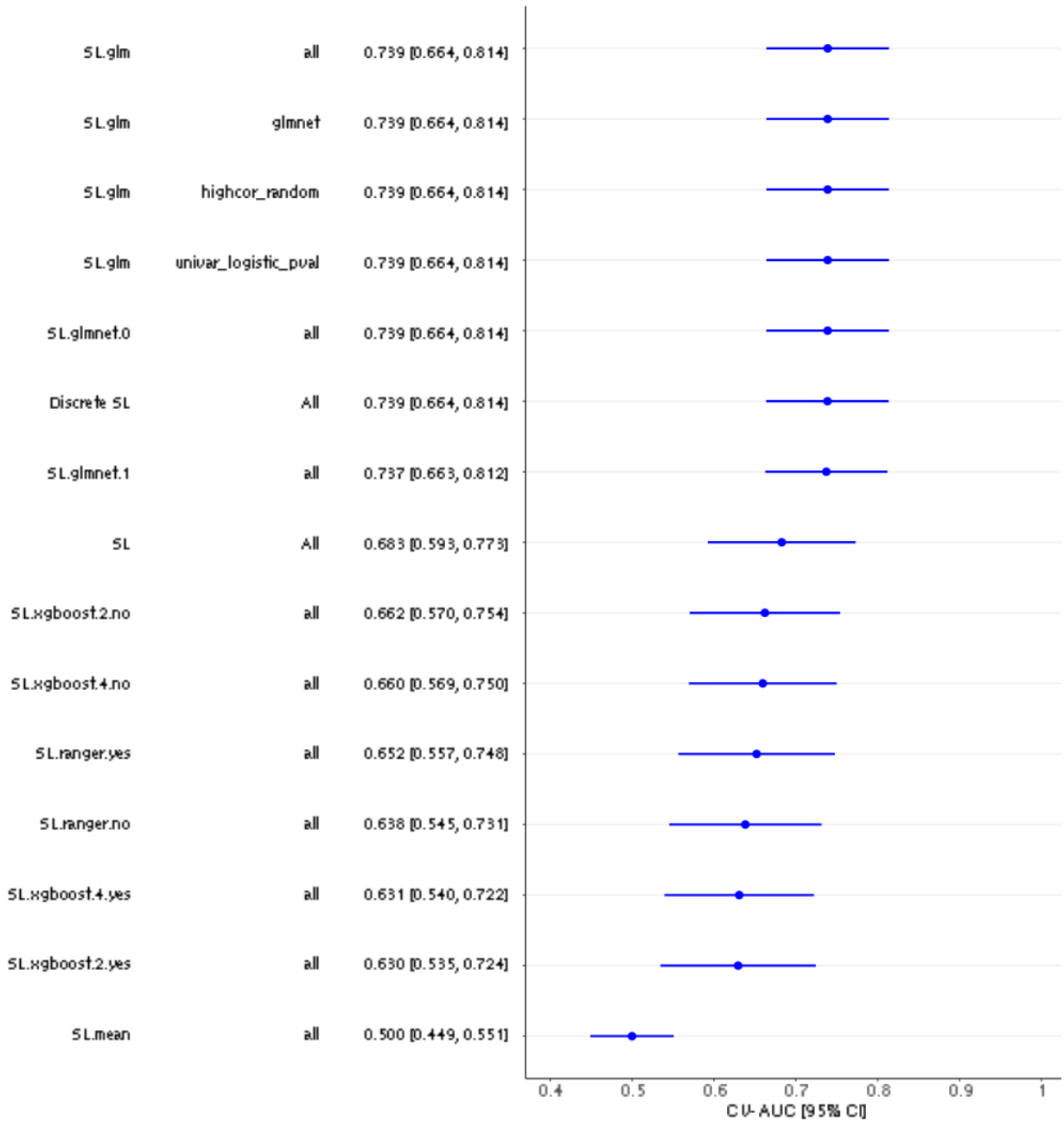

Figure 20: 7_varset_lnabMN50: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

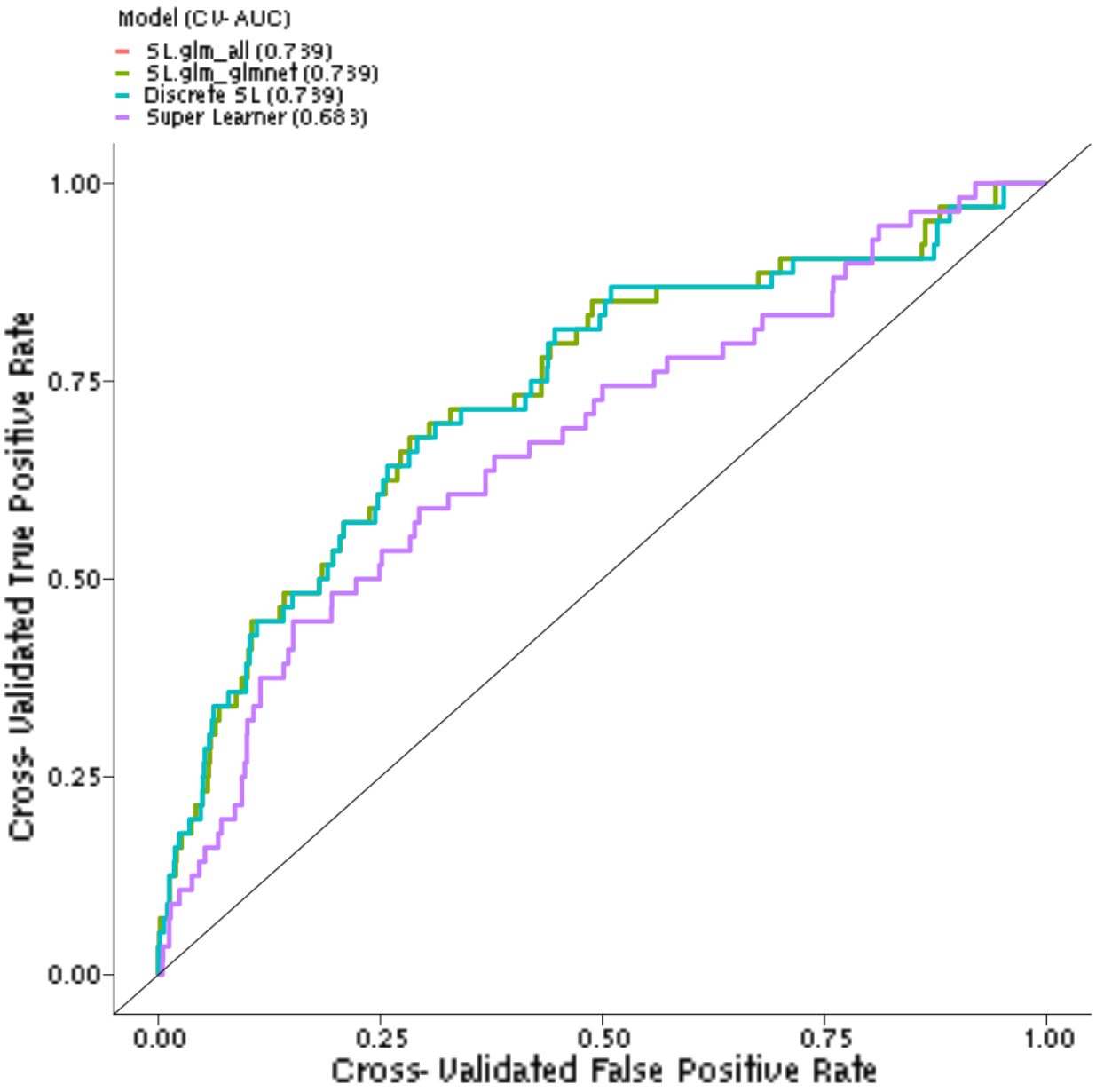

Figure 21: 7_varset_lnabMN50: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

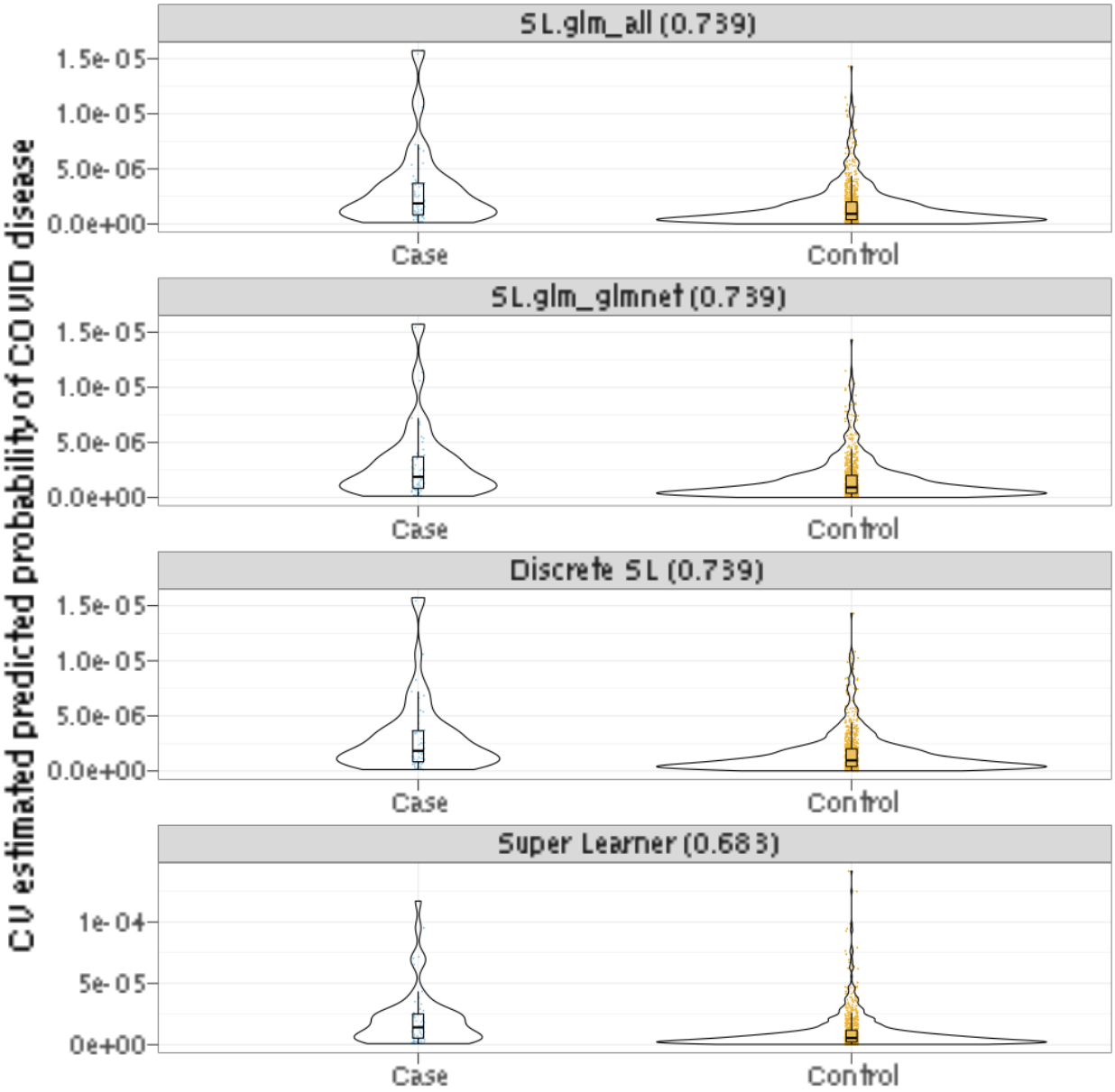

Figure 22: 7_varset_lnabMN50: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

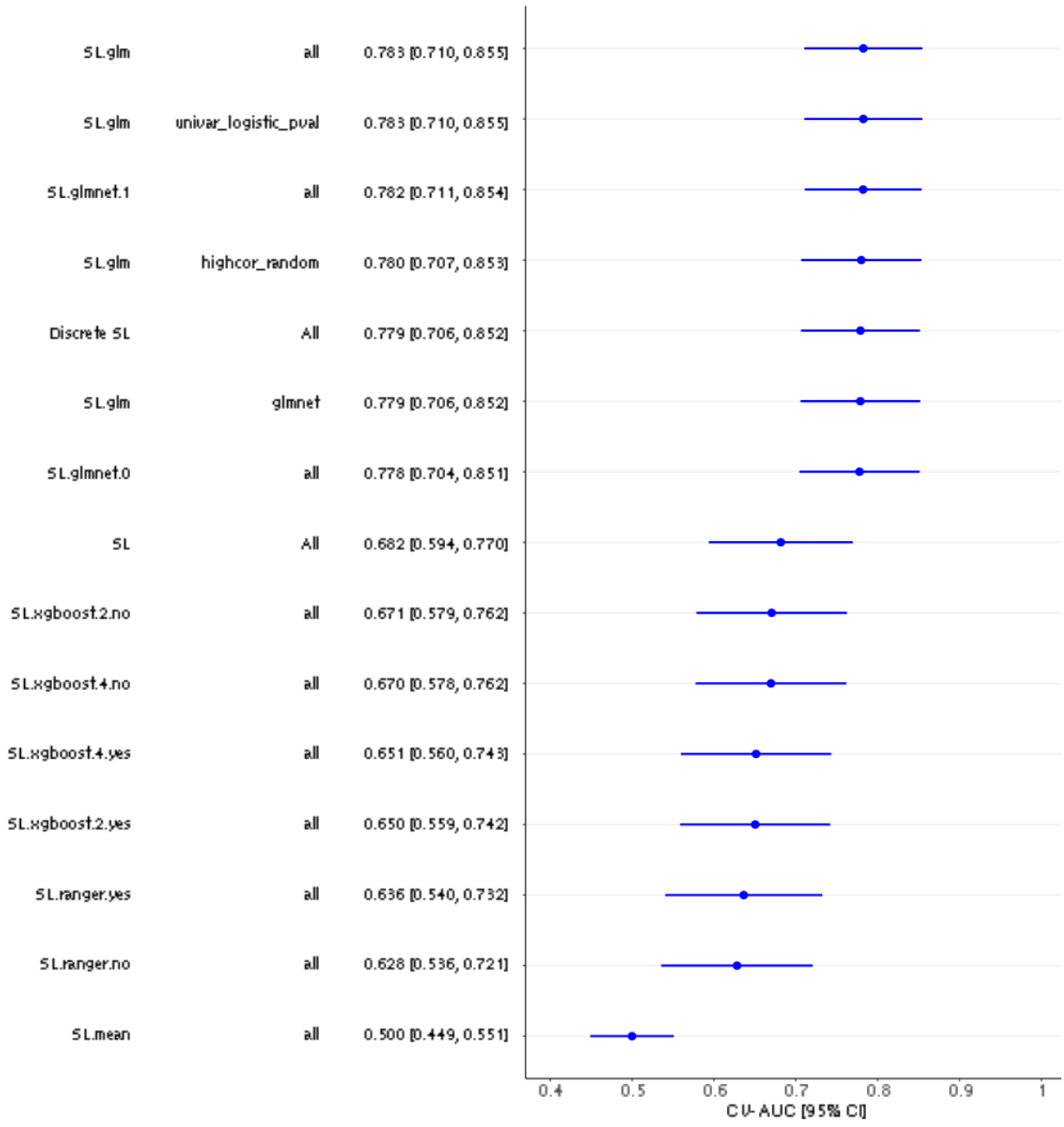

Figure 23: 12_varset_allMarkers: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

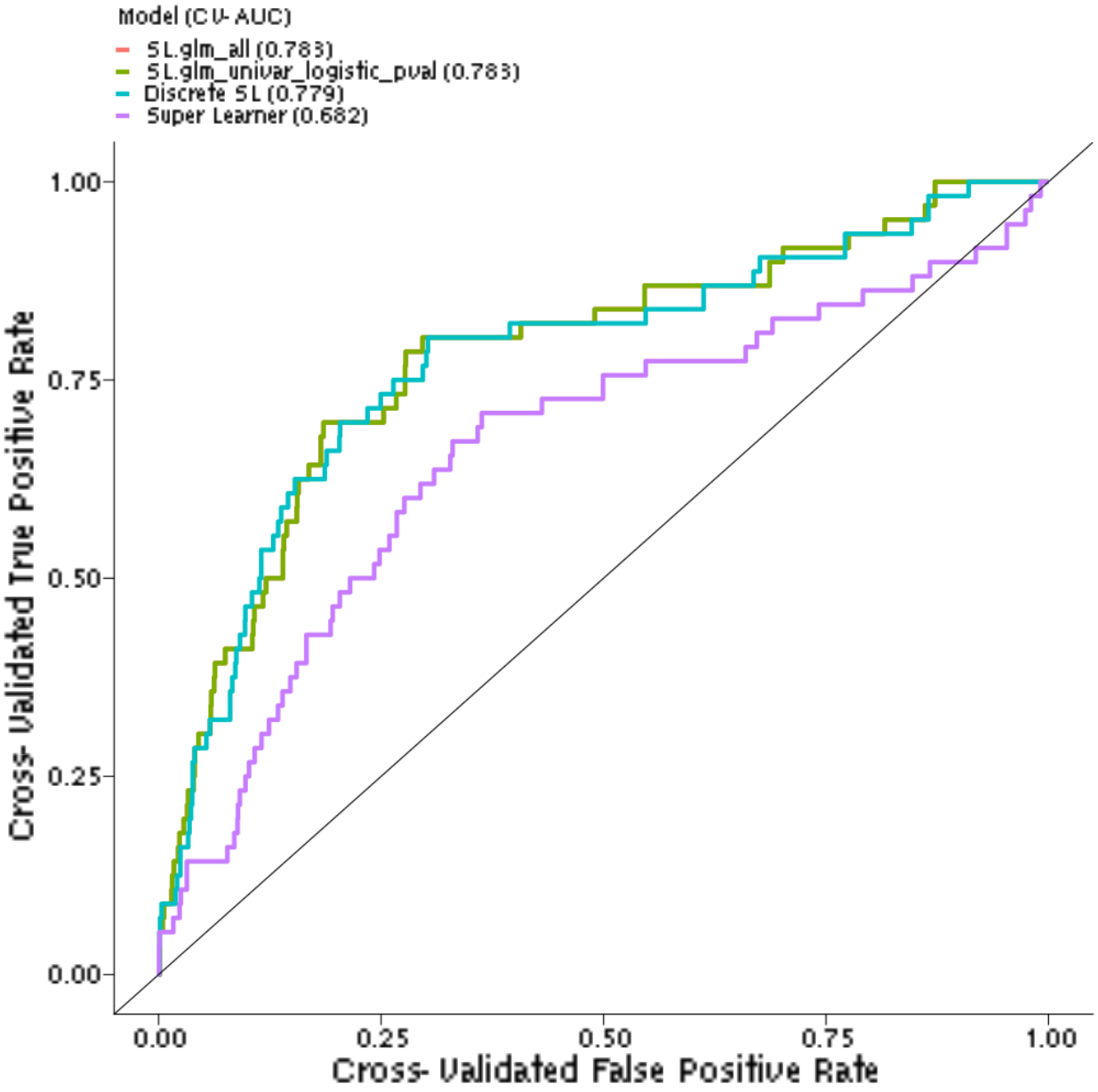

Figure 24: 12_varset_allMarkers: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

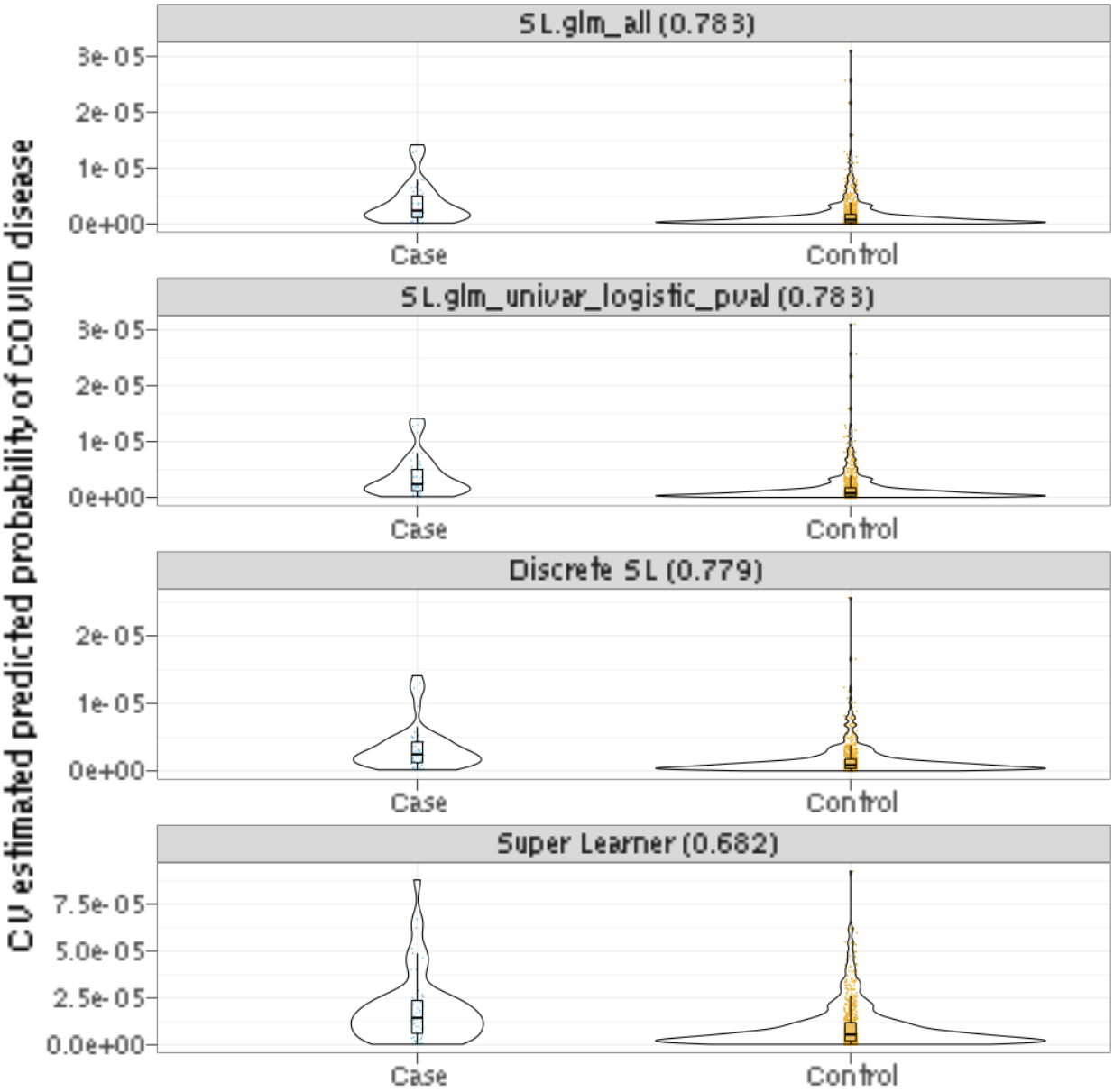

Figure 25: 12_varset_allMarkers: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

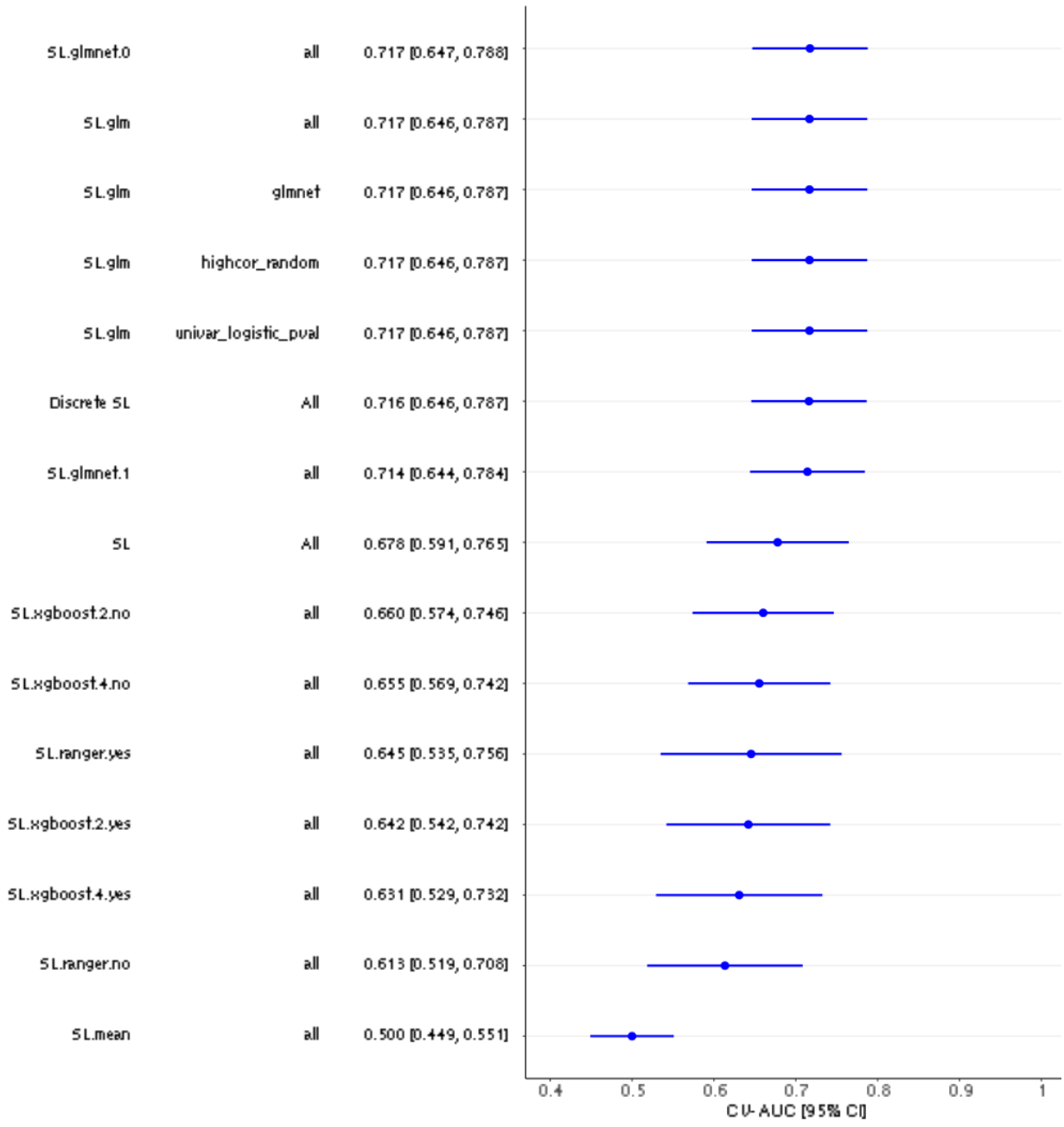

Figure 26: 2_baselineRiskFactors: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

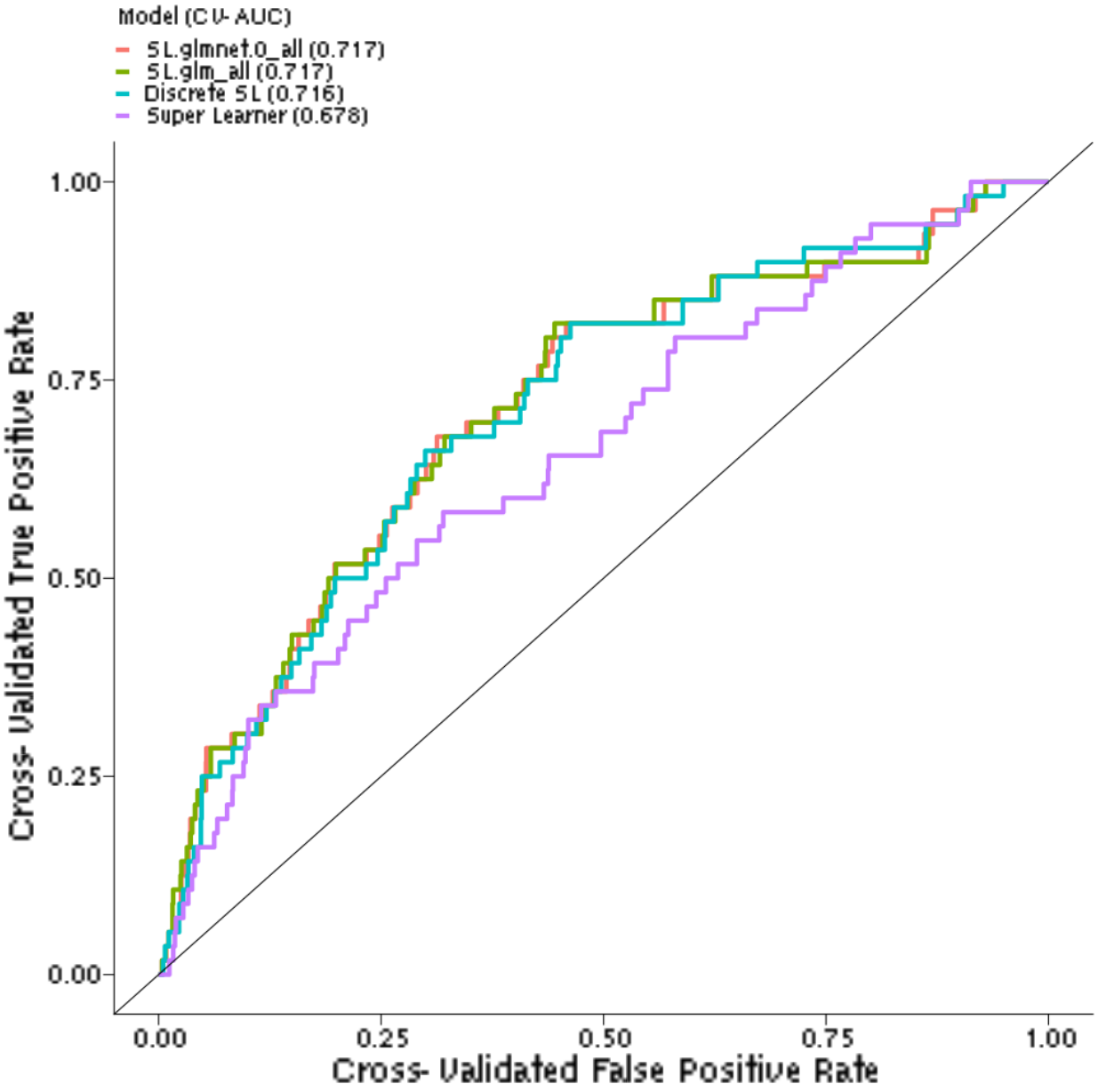

Figure 27: 2_baselineRiskFactors: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

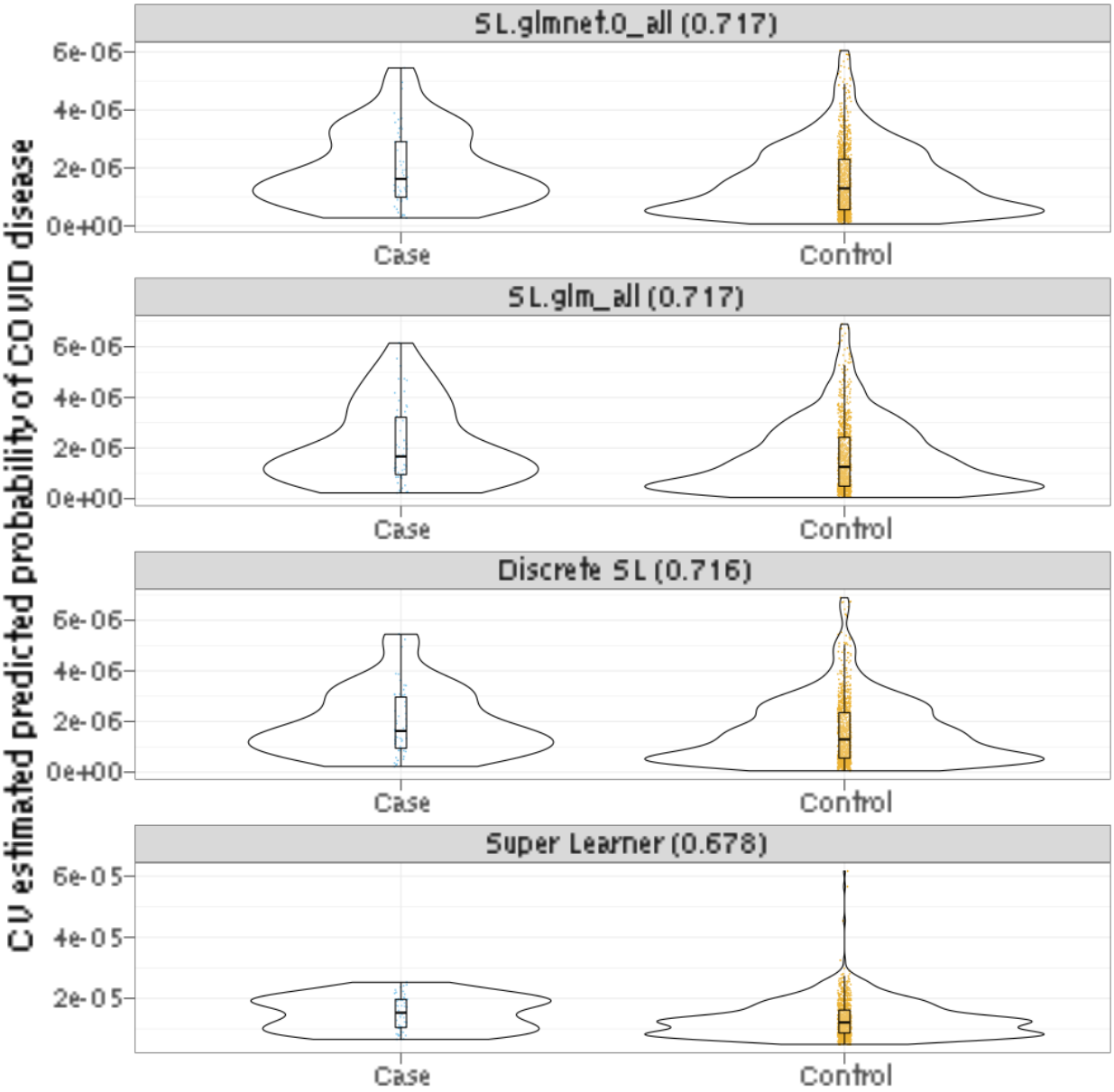

Figure 28: 2_baselineRiskFactors: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

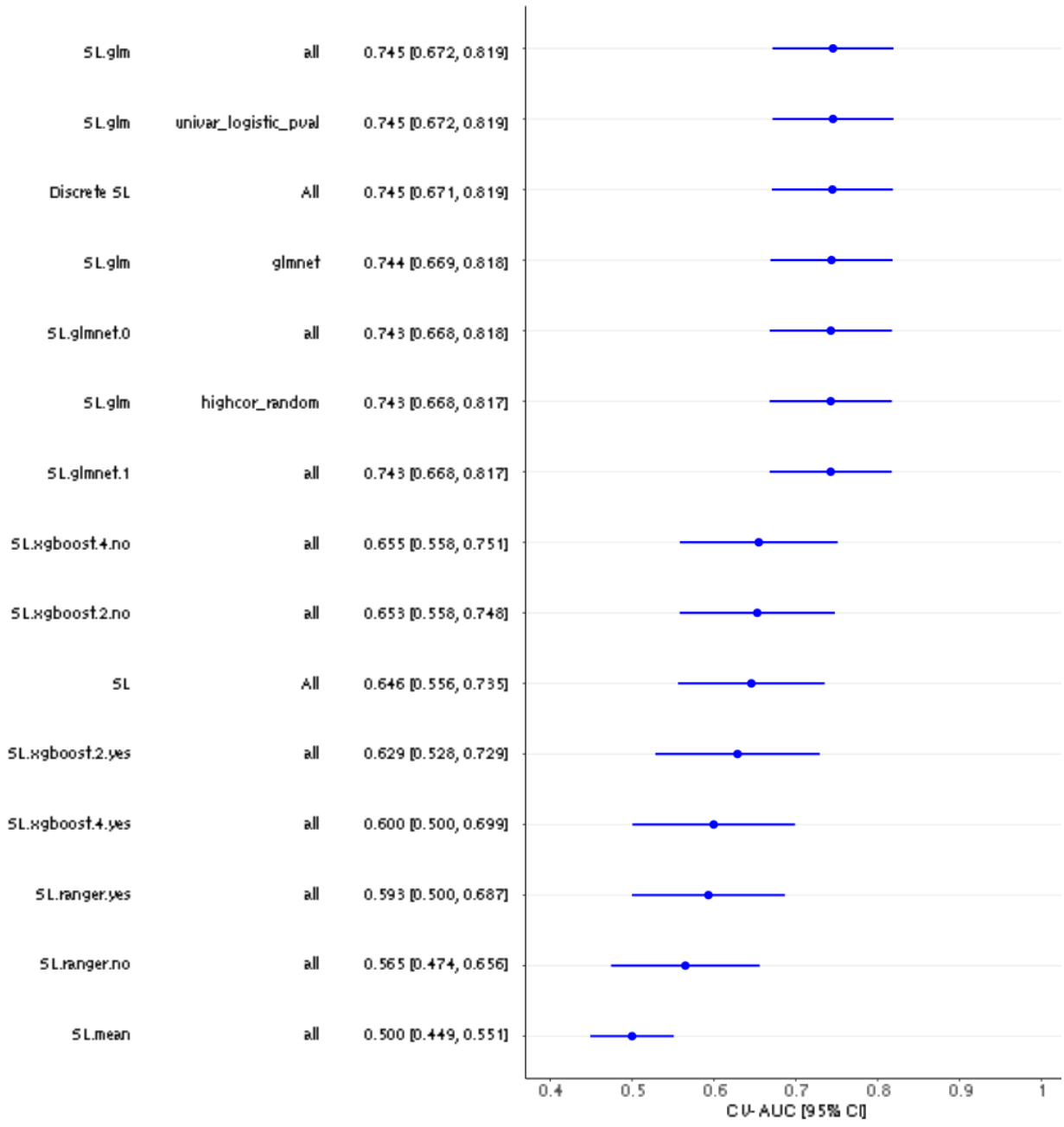

Figure 29: 5_varset_pnabID50: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

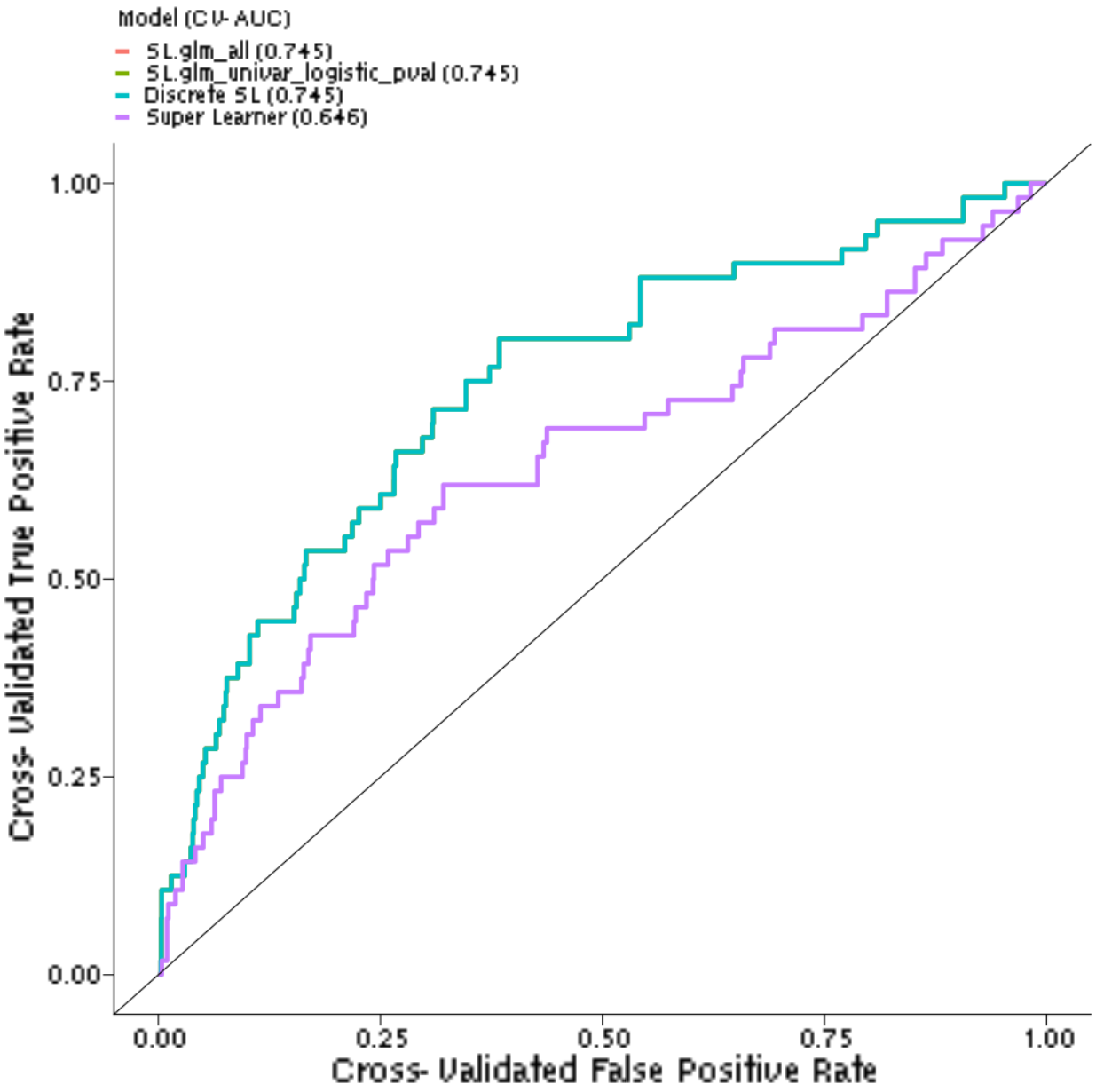

Figure 30: 5_varset_pnabID50: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

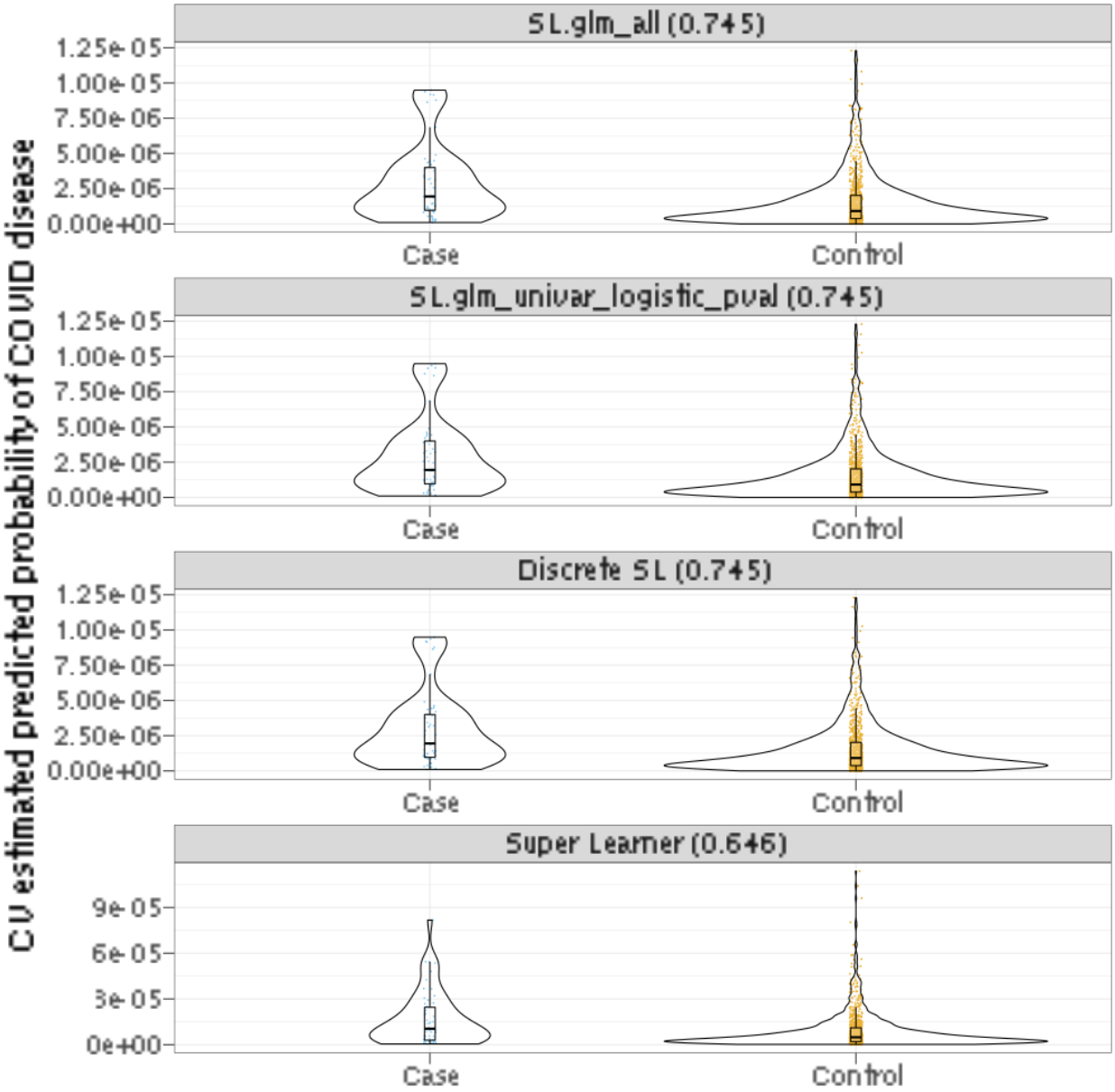

Figure 31: 5_varset_pnabID50: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

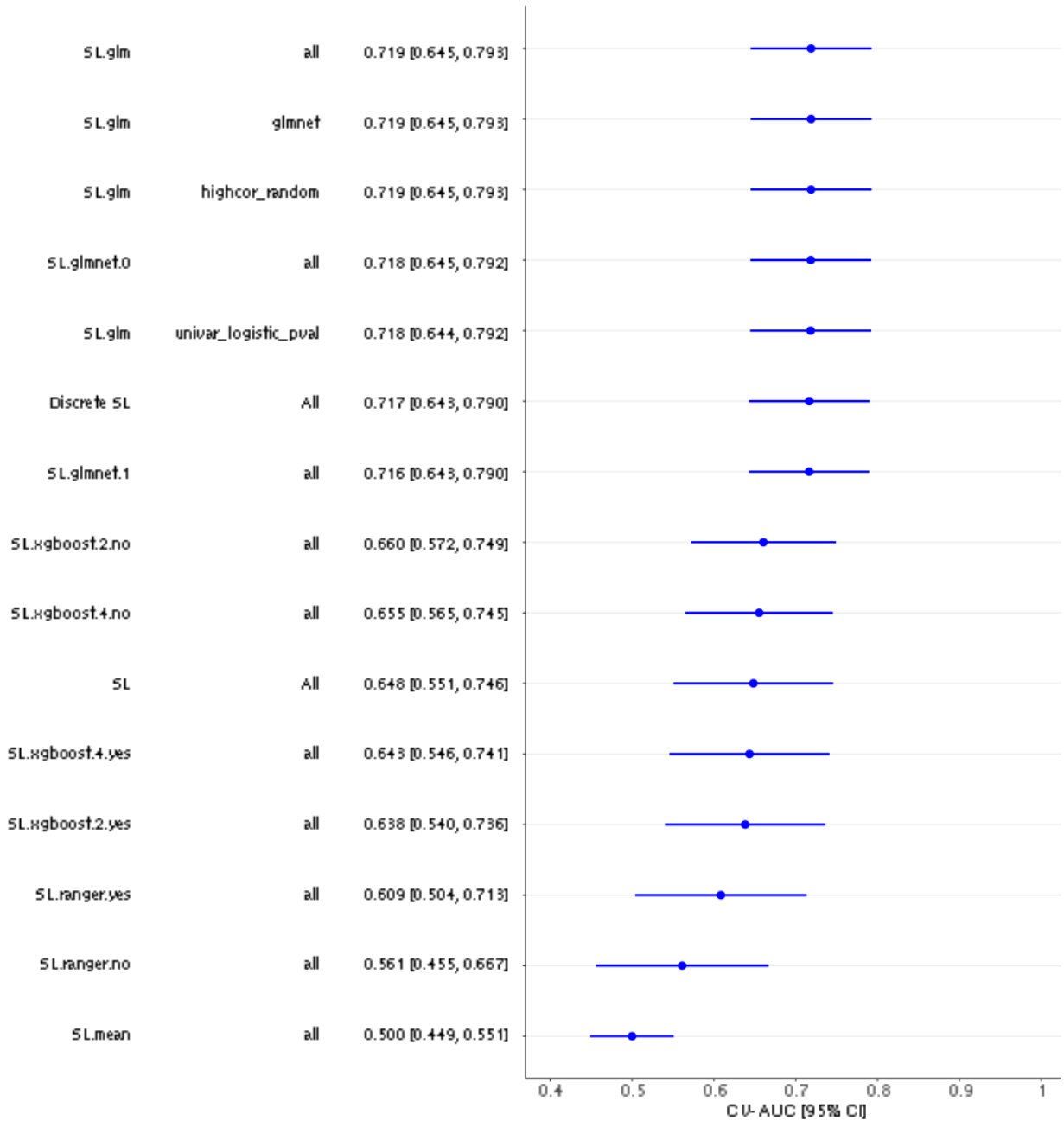

Figure 32: 4_varset_bAbRBD: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

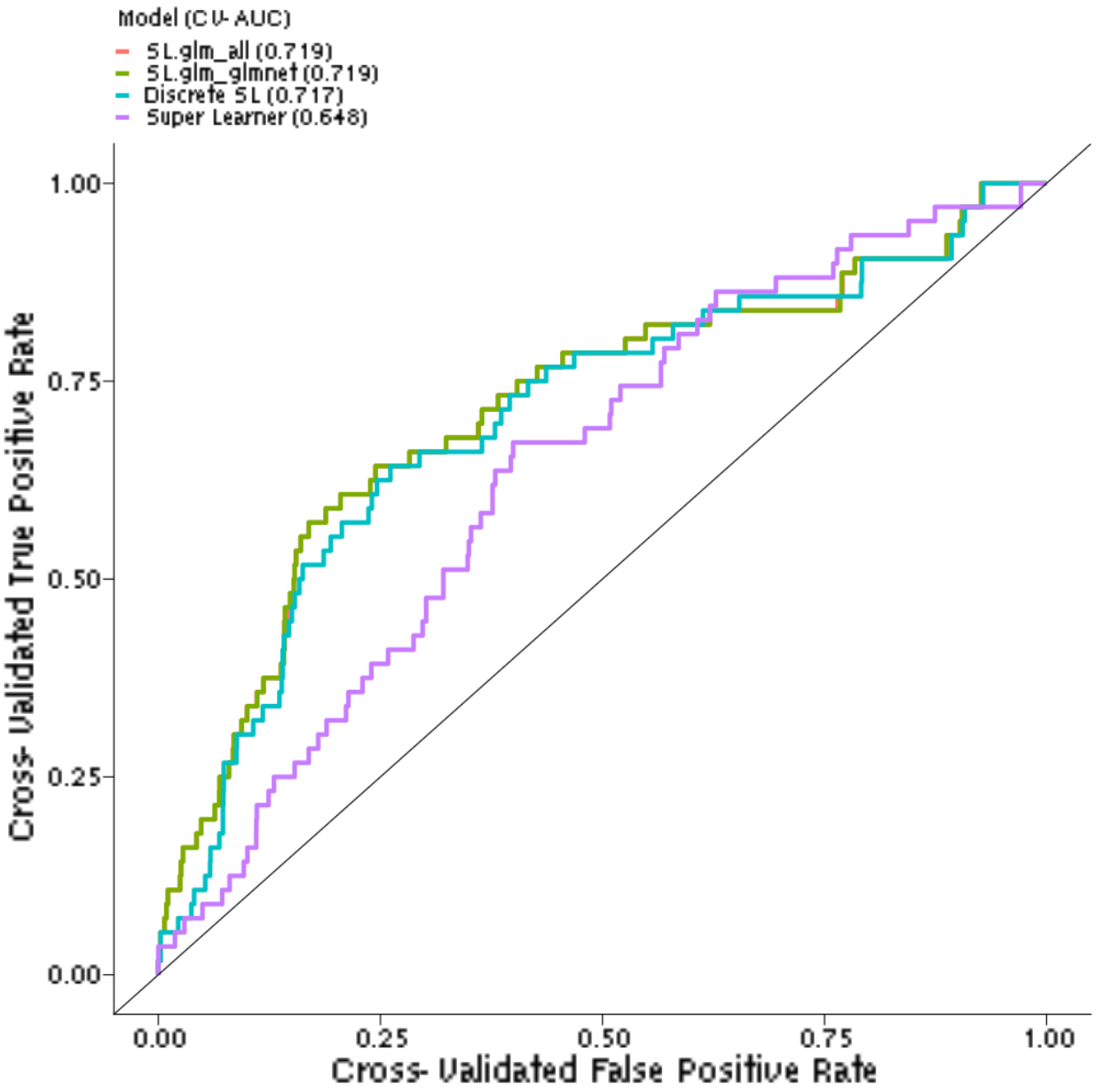

Figure 33: 4_varset_bAbRBD: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

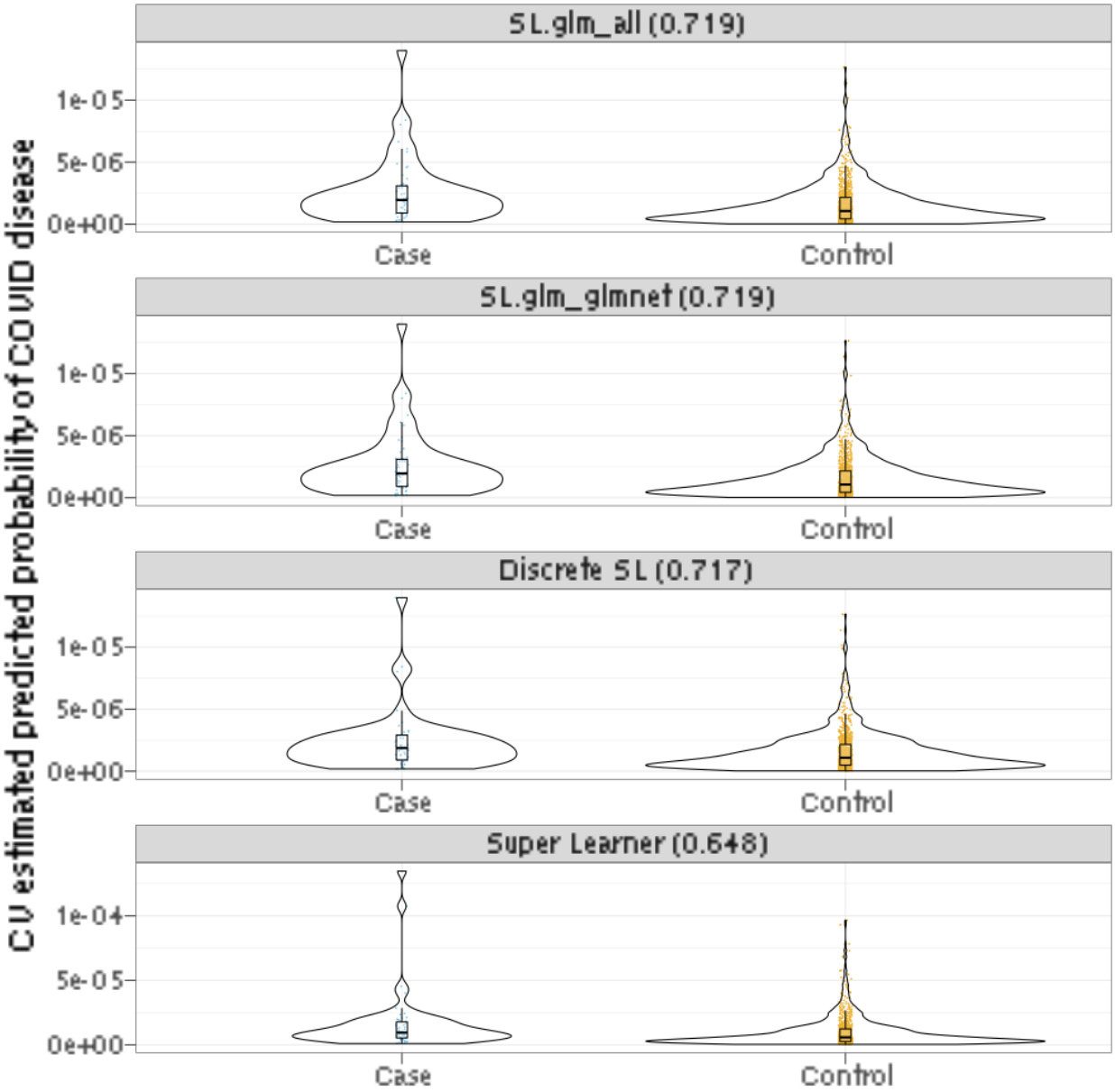

Figure 34: 4_varset_bAbRBD: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

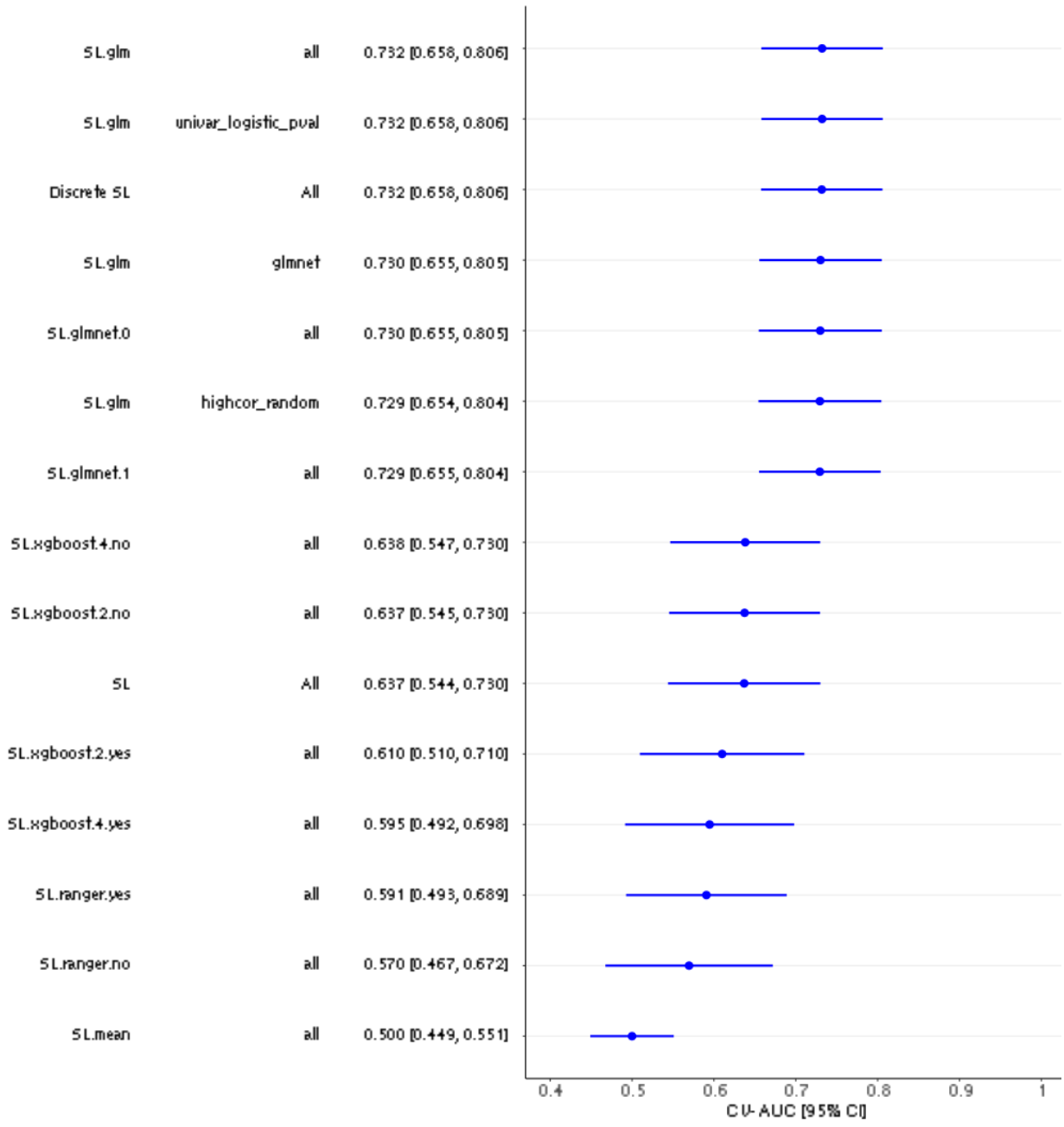

Figure 35: 6_varset_pnabID80: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

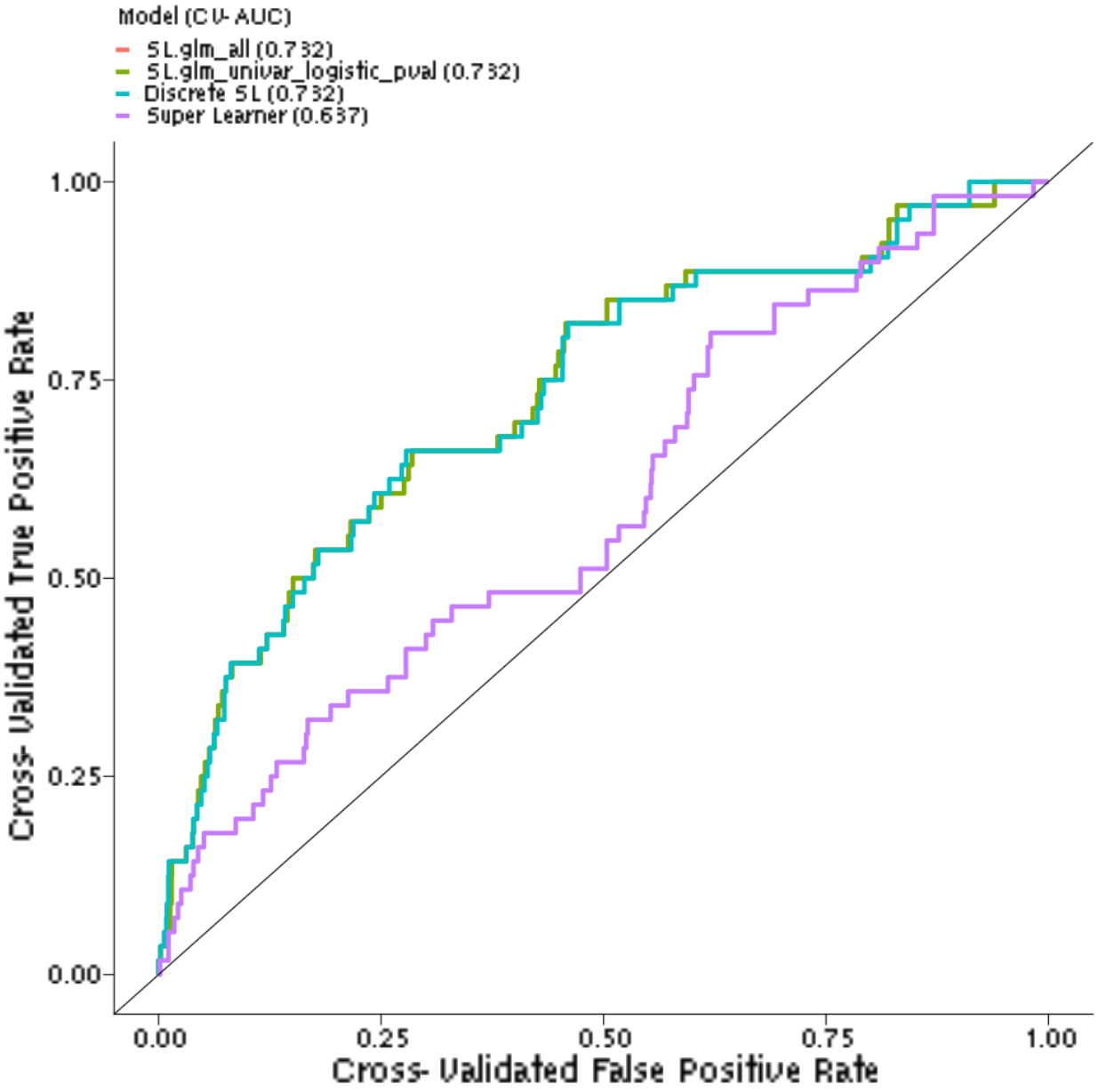

Figure 36: 6_varset_pnabID80: Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

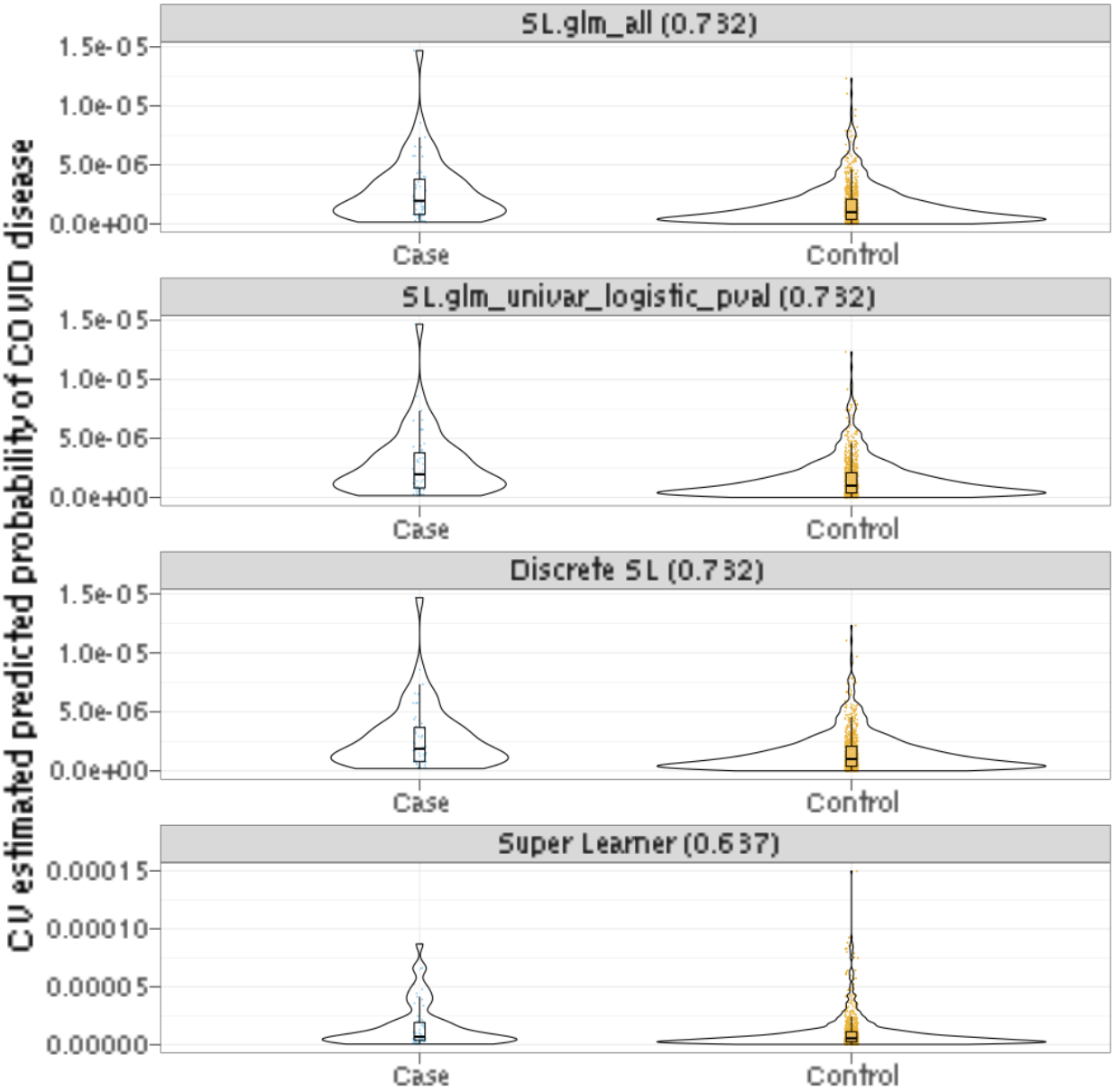

Figure 37: 6_varset_pnabID80: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.

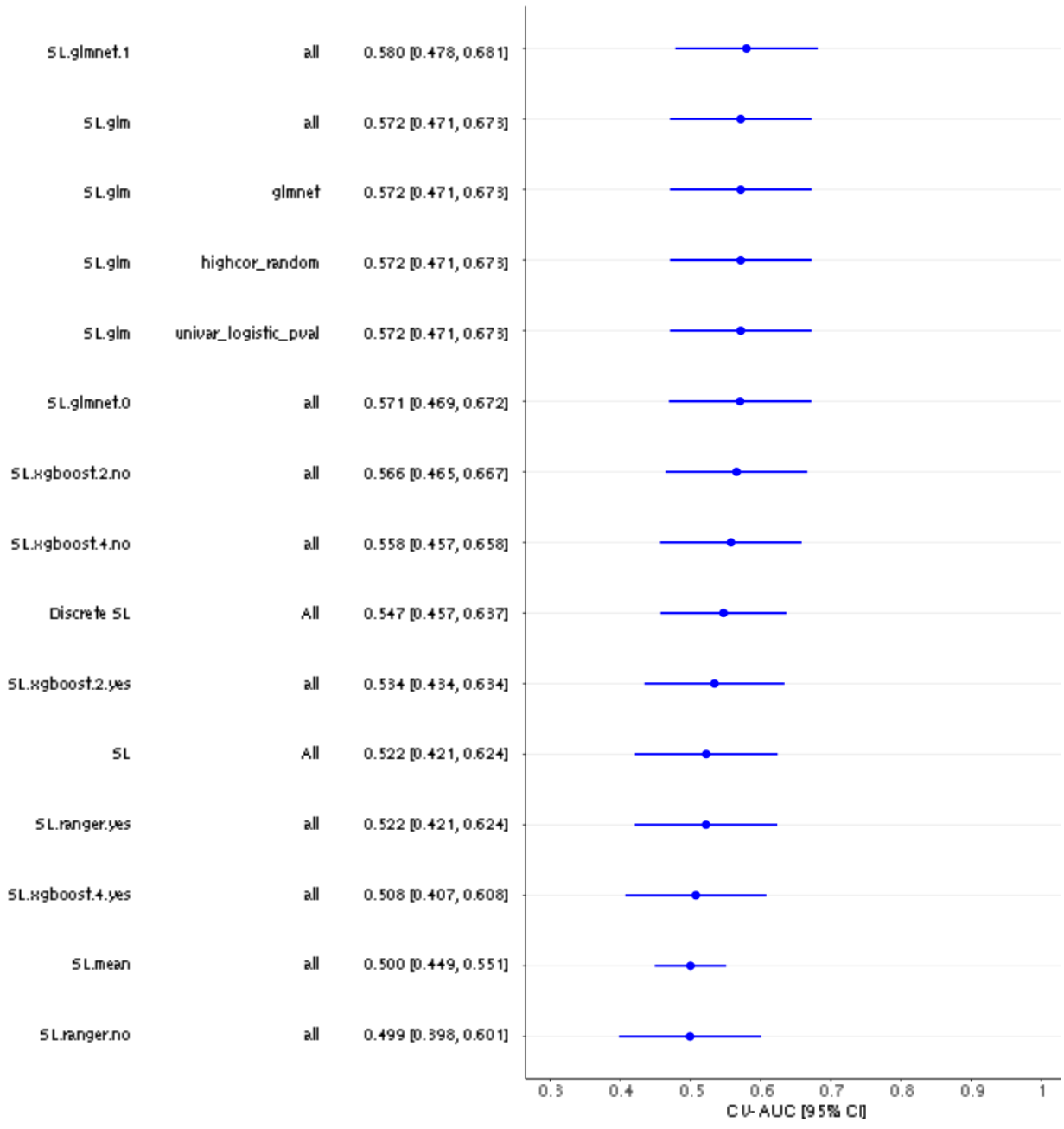

Figure 38: 1_noisyVars: Weighted CV-AUC (95% CI) of algorithms for predicting COVID-19 disease status starting 7 days after Day 57.

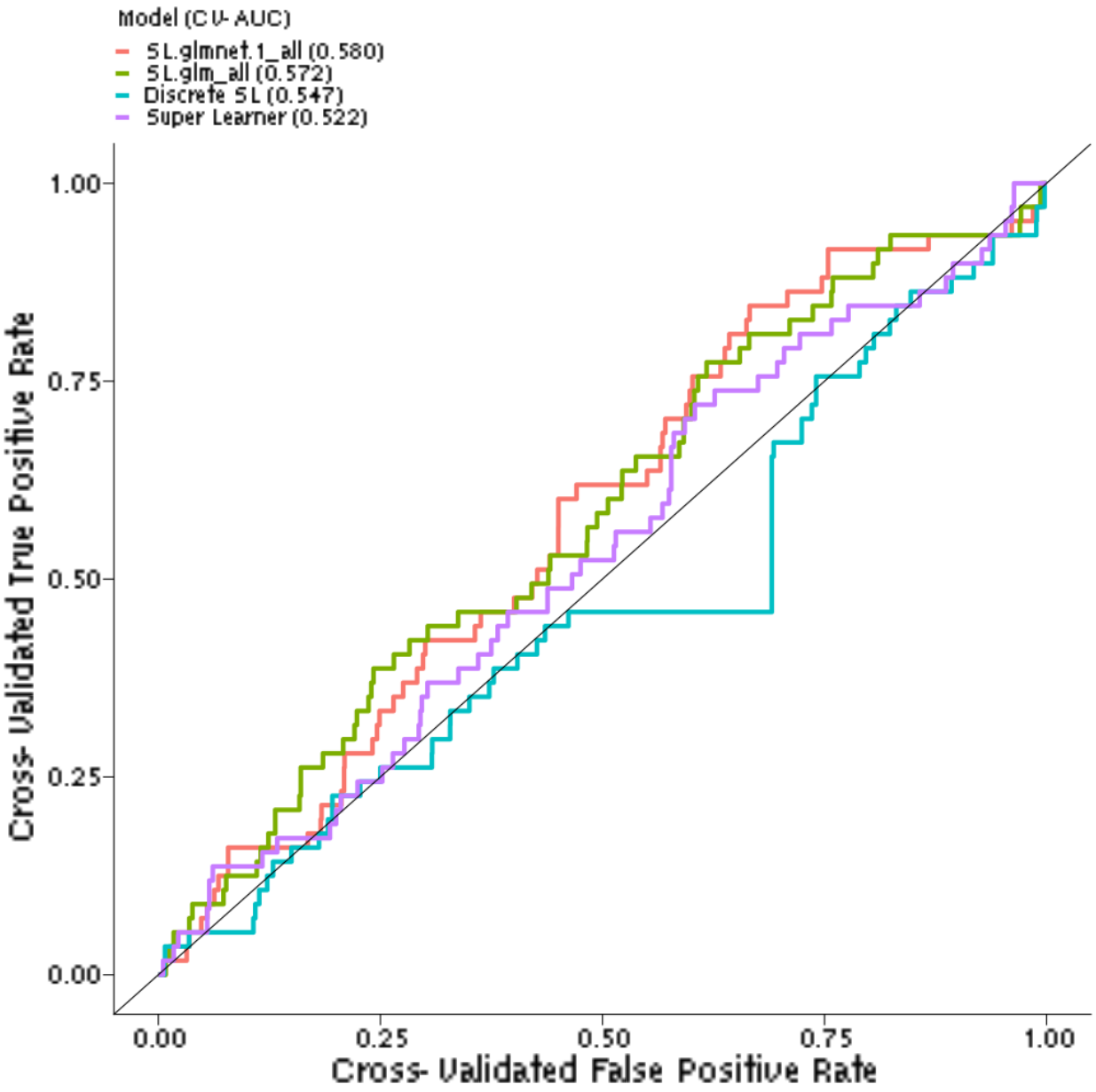

Figure 39: 1_noisyVars: Weighted Weighted CV-AUC ROC curves of top two individual learners along with Superlearner and discrete-SL.

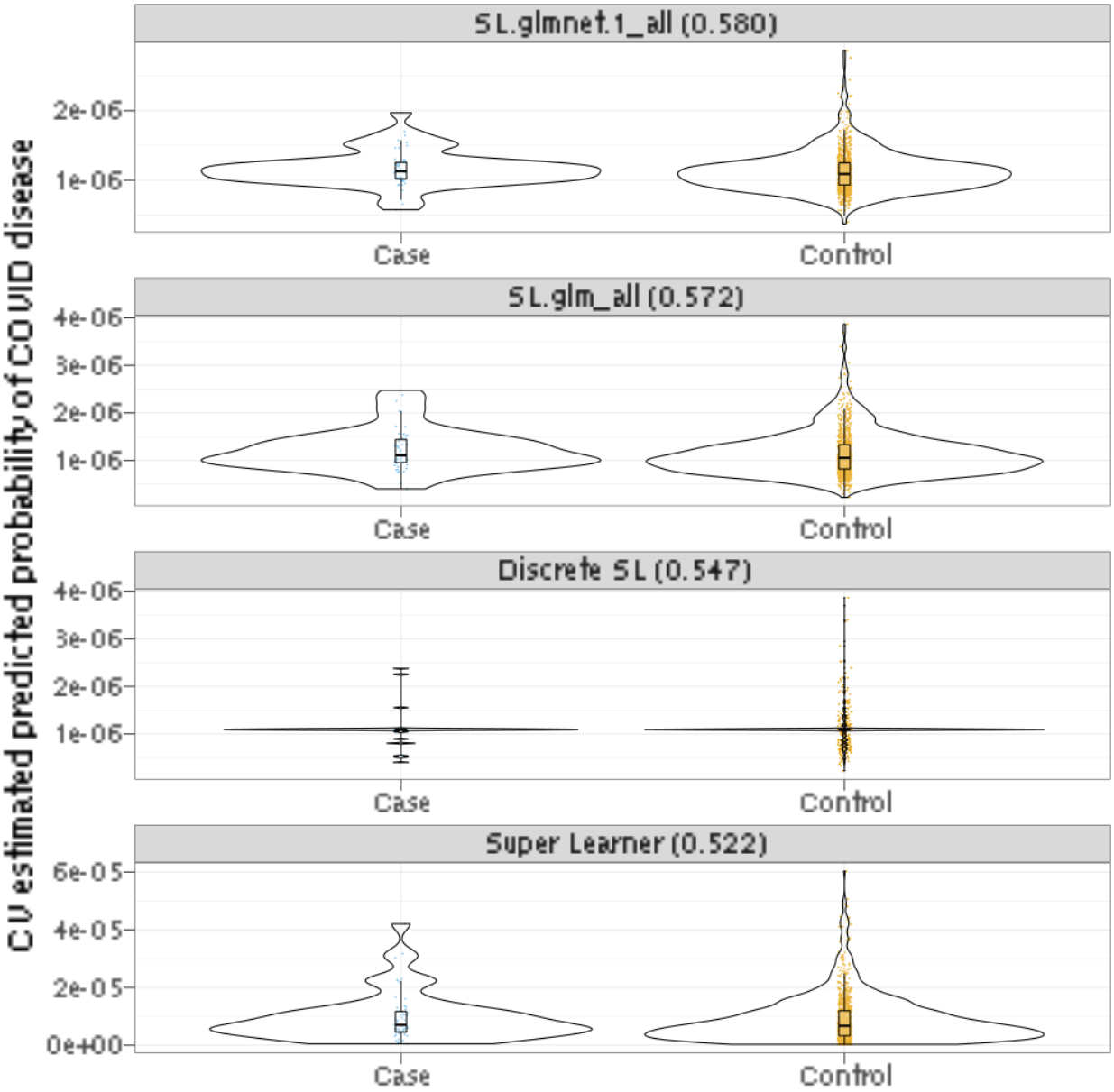

Figure 40: 1_noisyVars: Weighted prediction probability plots of top two individual learners along with Superlearner and discrete-SL.