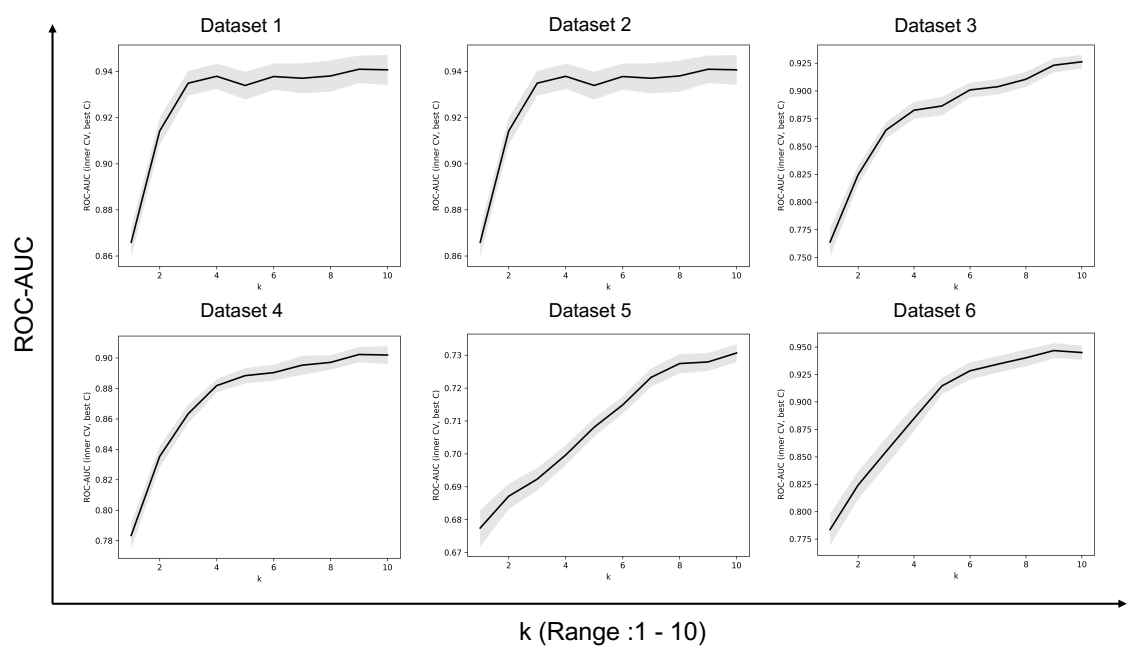
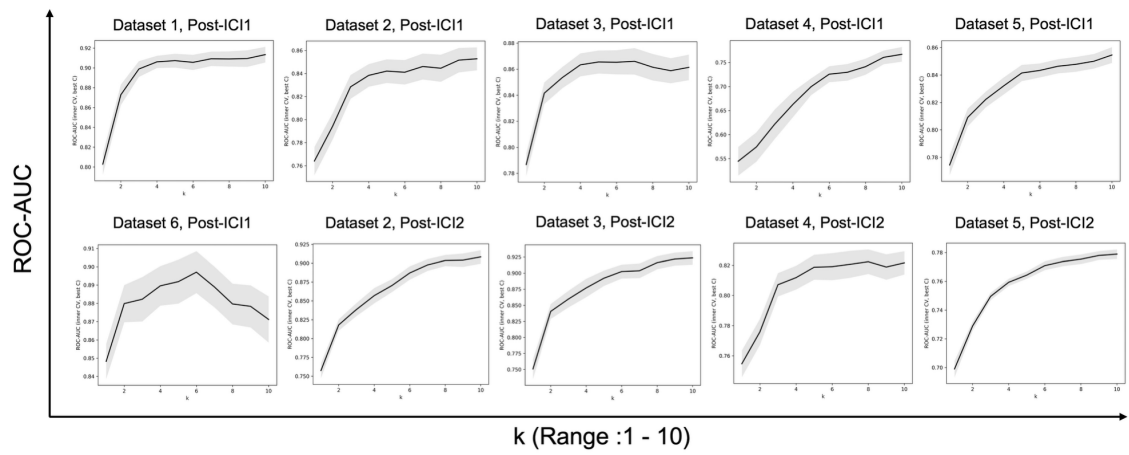


Supplementary Fig. S1. ROC-AUC across k-values in pre-ICI LC/MS datasets. Inner-CV ROC-AUC as a function of the number of top-ranked metabolites (k) in the six pre-ICI LC/MS datasets. Lines indicate the mean inner-CV ROC-AUC across outer-CV iterations, and shaded bands indicate ± 1 standard error across outer-CV iterations. ICI; immune checkpoint inhibitor, LC/MS; liquid chromatography/mass spectrometry, CV; cross-validation, ROC-AUC; area under the receiver operating characteristic curve



Supplementary Fig. S2. ROC-AUC across k-values in post-ICI LC/MS datasets. Inner-CV ROC-AUC as a function of the number of top-ranked metabolites (k) in the ten post-ICI LC/MS datasets. Lines indicate the mean inner-CV ROC-AUC across outer-CV iterations, and shaded bands indicate ± 1 standard error across outer-CV iterations. ICI; Immune checkpoint inhibitor, LC/MS; Liquid chromatography/mass spectrometry, CV; cross-validation, ROC-AUC; area under the receiver operating characteristic curve



Supplementary Table S1. Mean selected number of top-ranked features (k) in the pre-ICI datasets across validation strategies.

Dataset	k of nCV mean (min, max)	k of Enet-nest (fixed)	k of L2-nest mean (min, max)	k of non-nested CV (fixed)
1	6.68 (3, 10)	6	6.04 (4, 10)	6
2	5.24 (1, 9)	7	7.44 (1, 9)	7
3	4.28 (2, 9)	5	6.52 (2, 10)	5
4	7.40 (2, 10)	4	4.68 (4, 9)	4
5	7.00 (4, 10)	10	8.52 (7, 10)	10
6	6.16 (3, 9)	10	8.80 (6, 10)	10

Mean selected k within each pre-ICI dataset for nCV, Enet-nest, L2-nest, and non-nested CV, showing that the reduction in optimism was not simply explained by lower model complexity. ICI; immune checkpoint inhibitor, ROC-AUC; area under the receiver operating characteristic curve, CV; cross-validation