

Supplementary Material

Table S1 Threshold-dependent summaries at the fixed threshold of 0.5

Feature set	Sensitivity	Specificity
19 features	0.531	0.779
15 features	0.543	0.745
10 features	0.494	0.758

Table S2 Paired comparisons of feature sets: 19 features versus 15 and 10 features

Comparison	Metric	Difference	95% CI	Raw p	Holm-adjusted p
19 vs 15	ROC-AUC	0.014583	-0.020518 to 0.051599	0.408459	1.000000
19 vs 15	AP	0.042373	-0.003878 to 0.094359	0.075892	0.455354
19 vs 15	F1	0.007893	-0.070664 to 0.084338	0.857514	1.000000
19 vs 10	ROC-AUC	0.018560	-0.027458 to 0.064003	0.412559	1.000000
19 vs 10	AP	0.035014	-0.018927 to 0.092931	0.233977	1.000000
19 vs 10	F1	0.038217	-0.029984 to 0.106148	0.296870	1.000000

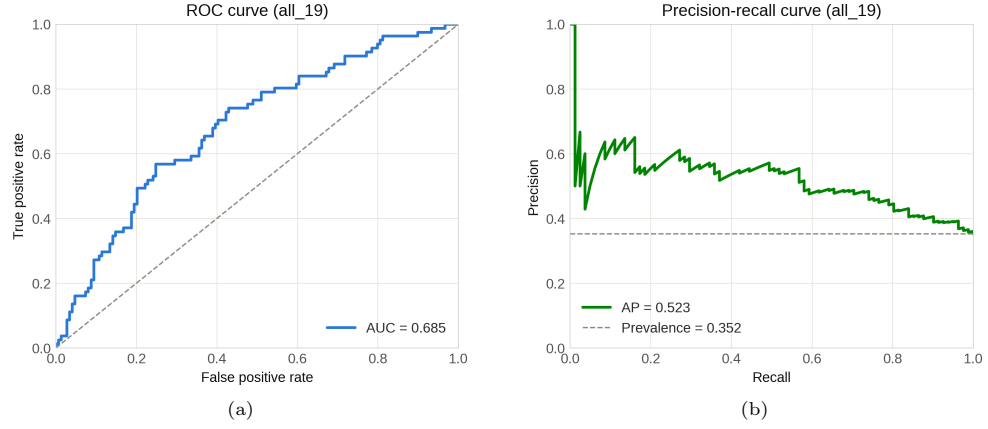


Figure S1 ROC curves (a) and precision–recall curves (b) from internal validation for the 19, 15 and 10-feature configurations. The area under each precision–recall curve corresponds to average precision (AP).

Table S3 Feature stability across outer folds

Feature	Mean rank	Mean importance	Positive-importance frequency	Availability
Tumor grade	1.84	0.119	1.00	25/25
pN	3.44	0.092	1.00	25/25
Vascular invasion	4.48	0.074	1.00	25/25
Solid pattern	4.92	0.066	1.00	25/25
Micropapillary pattern	6.24	0.056	1.00	25/25
Tumor size	7.44	0.047	1.00	25/25
Age	7.72	0.043	1.00	25/25
Lepidic pattern	8.44	0.041	0.96	25/25
Number of lymph nodes	8.56	0.040	1.00	25/25
Histologic type	9.4	0.034	1.00	25/25

Confusion matrix (all_19)			
Actual negative	116 50.4%	33 14.3%	149 (total)
Actual positive	38 16.5%	43 18.7%	81 (total)
Total	154 (total)	76 (total)	230 (total)
	Predicted negative	Predicted positive	Total

Figure S2 Confusion matrix for the 19-feature configuration at the fixed threshold of 0.5: TP = 43, FN = 38, FP = 33, and TN = 116.

Canonical aggregate point estimates with historical bootstrap CIs

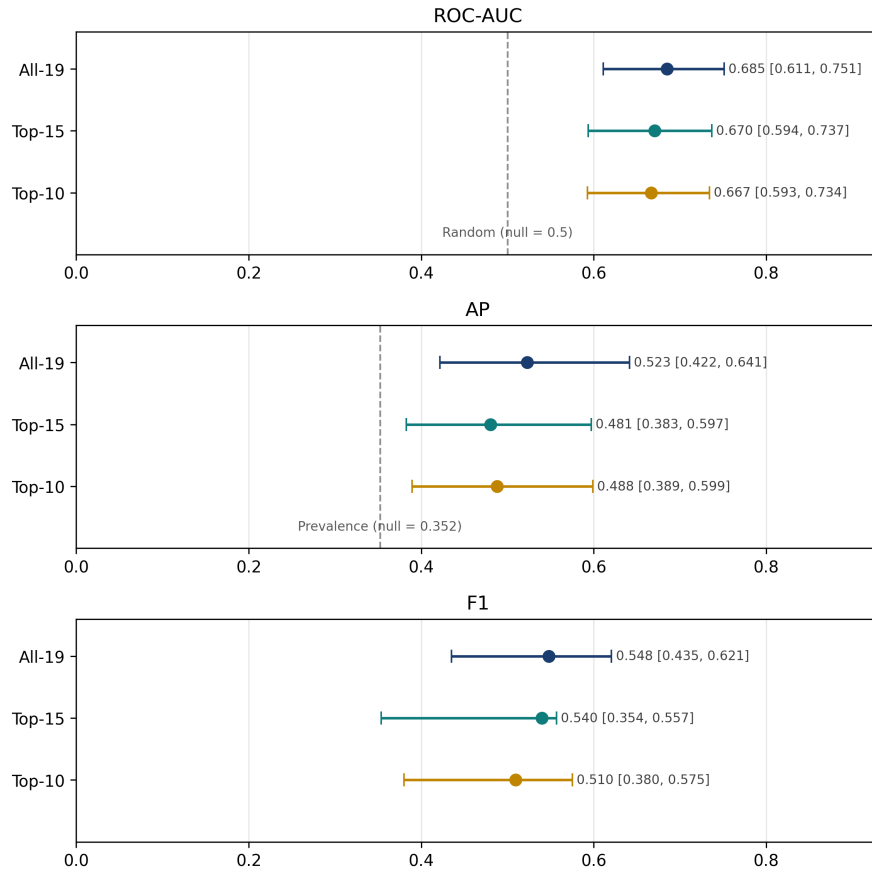


Figure S3 Primary and secondary permutation tests. ROC-AUC was evaluated against 0.5; AP was evaluated against the observed positive-class prevalence of 0.352.

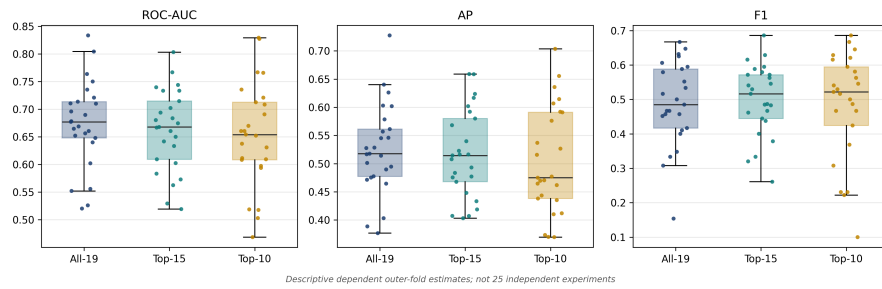


Figure S4 Variability of ROC-AUC and AP point estimates across the 25 outer folds. The estimates are dependent descriptive quantities.

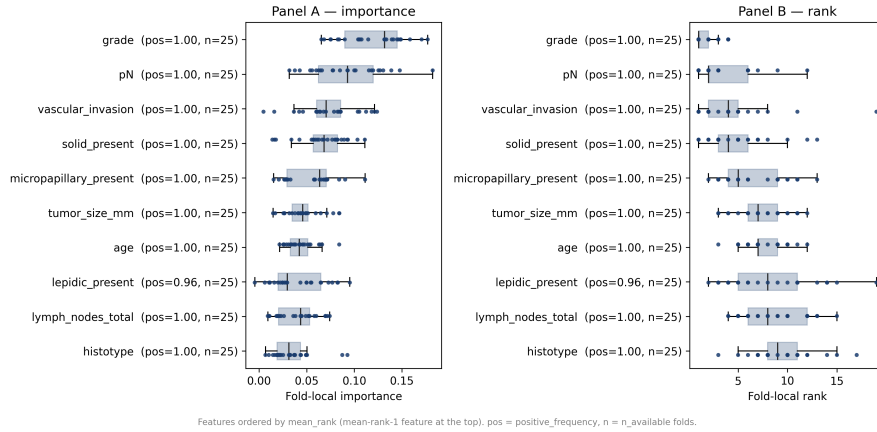


Figure S5 Ten features with the most stable positions by mean rank across outer test folds: tumor grade, pN, vascular invasion, solid pattern, micropapillary pattern, tumor size, age, lepidic pattern, number of lymph nodes, and histologic type.

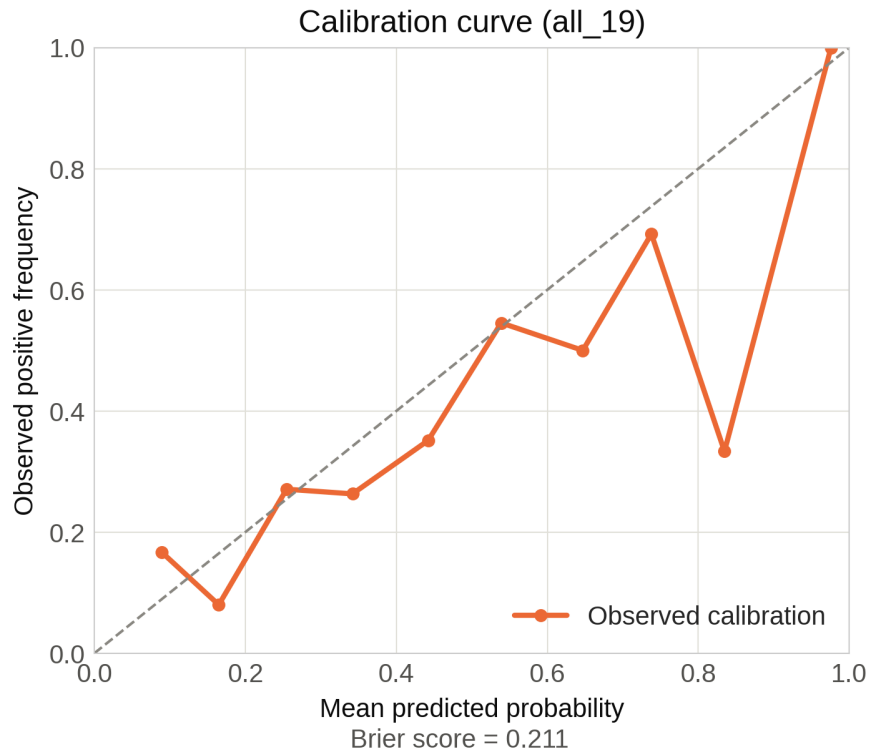


Figure S6 Calibration curve for the primary 19-feature configuration based on aggregated patient-level OOF predictions. The dashed line represents ideal calibration; Brier score = 0.211.