Supplementary Table 1 Missing data summary by variable and malignancy status

Variable	Total missing n (%)	Missing in Malignancy + (n = 23)	Missing in Malignancy − (n = 1,581)
Glucose (mg/dL)	10 (0.6%)	0 (0.0%)	10 (0.6%)
Triglyceride (mg/dL)	56 (3.5%)	1 (4.3%)	55 (3.5%)
HDL (mg/dL)	57 (3.6%)	1 (4.3%)	56 (3.5%)
LDL (mg/dL)	57 (3.6%)	1 (4.3%)	56 (3.5%)
AST (U/L)	8 (0.5%)	0 (0.0%)	8 (0.5%)
ALT (U/L)	8 (0.5%)	0 (0.0%)	8 (0.5%)
Neutrophil	4 (0.2%)	0 (0.0%)	4 (0.3%)
Lymphocyte	4 (0.2%)	0 (0.0%)	4 (0.3%)
Platelet	4 (0.2%)	0 (0.0%)	4 (0.3%)
TyG index	56 (3.5%)	1 (4.3%)	55 (3.5%)
AIP	57 (3.6%)	1 (4.3%)	56 (3.5%)
FIB-4	8 (0.5%)	0 (0.0%)	8 (0.5%)
NLR	4 (0.2%)	0 (0.0%)	4 (0.3%)
PLR	4 (0.2%)	0 (0.0%)	4 (0.3%)
SII	4 (0.2%)	0 (0.0%)	4 (0.3%)
NHR	57 (3.6%)	1 (4.3%)	56 (3.5%)

Age and sex had no missing values. Lipid panel parameters (triglyceride, HDL, LDL) and their derived indices (AIP, NHR, TyG) share a common missing pattern (n ≈ 56–57, 3.5–3.6%), likely reflecting patients who did not have a fasting lipid panel performed prior to colonoscopy. Missing values were addressed by median imputation fitted within each training split during cross-validation.

Supplementary Table 2 Pre-colonoscopy laboratory parameters and derived indices, stratified by malignancy status

Variable	Overall (N = 1,604)	Malignancy + (n = 23)	Malignancy − (n = 1,581)	p *	Missing n (%)
Raw laboratory parameters					
Glucose (mg/dL)	94.0 (86.0–106.0)	104.0 (97.0–135.0)	93.0 (86.0–106.0)	**< 0.001**	10 (0.6%)
Triglyceride (mg/dL)	127.0 (96.0–175.3)	137.0 (108.3–193.0)	126.0 (95.3–175.0)	0.248	56 (3.5%)
HDL (mg/dL)	49.0 (42.0–57.0)	44.0 (34.8–58.0)	49.0 (42.0–57.0)	0.275	57 (3.6%)
LDL (mg/dL)	110.0 (88.5–134.0)	114.0 (81.8–144.3)	110.0 (89.0–134.0)	0.775	57 (3.6%)

AST (U/L)	18.0 (15.0–22.0)	19.0 (11.5–21.5)	18.0 (15.0–22.0)	0.393	8 (0.5%)
ALT (U/L)	16.0 (12.0–21.0)	13.0 (12.0–17.0)	16.0 (12.0–21.0)	0.187	8 (0.5%)
Neutrophil (10^3/μL)	4145 (3278–5240)	4400 (3245–6455)	4140 (3280–5240)	0.182	4 (0.2%)
Lymphocyte (10^3/μL)	2155 (1740–2600)	1690 (1340–2370)	2160 (1750–2610)	**0.013**	4 (0.2%)
Platelet (10^3/μL)	270.5 (224.8–318.0)	329.0 (284.5–404.5)	270.0 (224.0–318.0)	**0.007**	4 (0.2%)
Derived inflammatory / metabolic indices					
TyG index	8.68 (8.38–9.10)	8.86 (8.56–9.32)	8.68 (8.38–9.10)	0.063	56 (3.5%)
AIP	0.42 (0.25–0.59)	0.47 (0.34–0.76)	0.42 (0.25–0.59)	0.153	57 (3.6%)
FIB-4	1.01 (0.74–1.37)	0.91 (0.60–1.17)	1.01 (0.74–1.37)	0.304	8 (0.5%)
NLR	1.90 (1.45–2.52)	2.69 (2.24–3.69)	1.89 (1.44–2.49)	**< 0.001**	4 (0.2%)
PLR	124.9 (99.0–161.2)	178.1 (127.3–264.2)	124.5 (98.7–160.5)	**< 0.001**	4 (0.2%)
SII	508.8 (369.1–716.9)	769.0 (577.8–1349.3)	505.3 (367.9–713.0)	**< 0.001**	4 (0.2%)
NHR	85.4 (62.1–115.5)	99.5 (62.2–160.1)	85.3 (62.2–115.0)	0.103	57 (3.6%)

*Data are presented as median (IQR). * Mann–Whitney U test. Bold p-values indicate statistical significance ($p < 0.05$). Missing values were handled by median imputation during model training.*

Supplementary Table 3 Post-test probability of malignancy across five pre-test prevalence scenarios, derived from the Youden-threshold likelihood ratios (Fagan's nomogram; Bayes' rule)

All features model (LR+ = 3.66, LR− = 0.514)

Pre-test prevalence	Post-test probability if test POSITIVE	Post-test probability if test NEGATIVE	Relative risk reduction if NEGATIVE
1.43 %	5.04 %	0.740 %	48.2 %
3.00 %	10.17 %	1.565 %	47.8 %
5.00 %	16.15 %	2.634 %	47.3 %
10.00 %	28.91 %	5.403 %	46.0 %
20.00 %	47.78 %	11.387 %	43.1 %

SHAP10 model (LR+ = 1.84, LR− = 0.173) — primary rule-out model

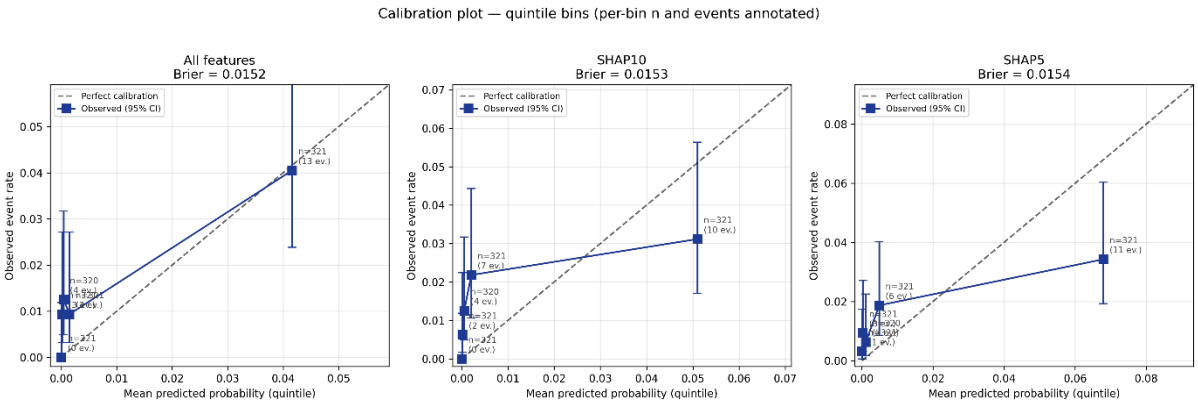
Pre-test prevalence	Post-test probability if test POSITIVE	Post-test probability if test NEGATIVE	Relative risk reduction if NEGATIVE
1.43 %	2.60 %	0.250 %	82.5 %

3.00 %	5.38 %	0.532 %	82.3 %
5.00 %	8.83 %	0.902 %	82.0 %
10.00 %	16.97 %	1.886 %	81.1 %
20.00 %	31.51 %	4.146 %	79.3 %

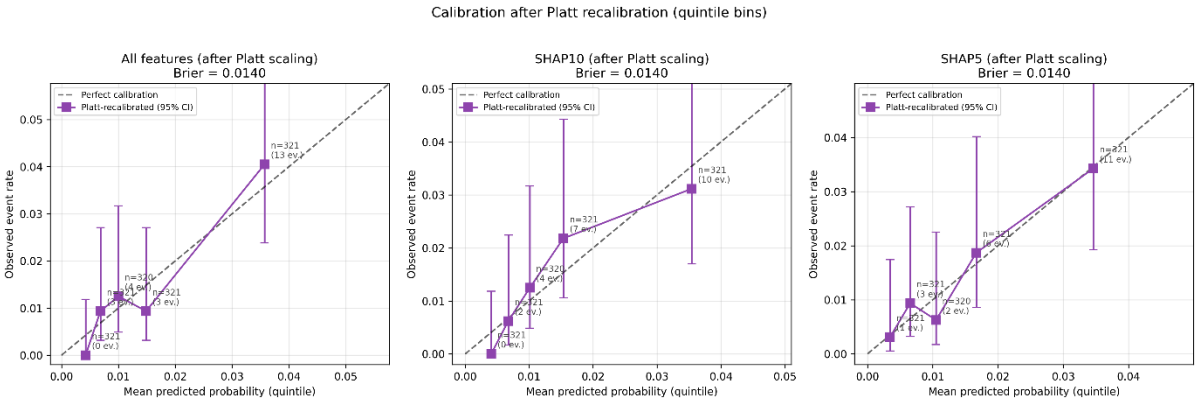
SHAP5 model (LR+ = 1.99, LR− = 0.415)

Pre-test prevalence	Post-test probability if test POSITIVE	Post-test probability if test NEGATIVE	Relative risk reduction if NEGATIVE
1.43 %	2.81 %	0.598 %	58.1 %
3.00 %	5.80 %	1.267 %	57.8 %
5.00 %	9.48 %	2.138 %	57.2 %
10.00 %	18.11 %	4.408 %	55.9 %
20.00 %	33.22 %	9.400 %	53.0 %

Post-test probabilities were calculated from the Youden-threshold likelihood ratios (Table 4) using Bayes' rule: post-test odds = pre-test odds × LR, where odds = probability / (1 − probability). The pre-test prevalence of 1.43 % corresponds to the observed malignancy rate in our cohort; 3 %, 5 %, 10 % and 20 % represent plausible pre-test probabilities for higher-risk referral populations (e.g. symptomatic patients, patients with positive FIT, patients with a first-degree family history of colorectal cancer, and surveillance cohorts respectively). Even under a 14-fold higher pre-test prevalence (20 %), the SHAP10 model retains a residual post-negative probability below 4.2 %, demonstrating that its rule-out performance is not an artefact of the baseline prevalence.



Supplementary Figure 1 Calibration plot-quintile bins.



Supplementary Figure 2 Calibration after Platt recalibration.