

Supplementary Table 1 Boosted trees specifications

Measure	Definition
Entropy <i>R</i> -square	$1 - \text{Loglike}(\text{model})/\text{Loglike}(0)$
Generalized <i>R</i> -square	$\{1 - [L(0)/L(\text{model})]^{(2/n)}\} / [1 - L(0)^{(2/n)}]$
Mean-Log p	$\sum -\text{Log}[\rho(j)]/n$
Root-mean-squared-error	$\sqrt{\sum [y(j) - \rho(j)]^2/n}$
Mean absolute deviation	$\sum y(j) - \rho(j) /n$
Misclassification rate	$\sum [\rho(j) \neq \rho_{\text{Max}}]/n$

Supplementary Table 2 Boosted trees configuration

Configuration	Set value
Number of layers	200
Splits per tree	3
Learning rate	0.1
Overfit penalty	0.0001
Minimum size split	5
Row sampling rate	1
Column sampling rate	1

Supplementary Table 3 Boosted tree model used for the prediction of colorectal cancer

Overall statistics

	Measure	Training
Entropy	<i>R</i> -square	0.9977
Generalized	<i>R</i> -square	0.9982
Mean-Log	p	0.0005
Root-mean-squared-error		0.0012
Mean absolute deviation		0.0005
Misclassification rate		0.0000
<i>n</i>		184

Prediction

	Actual	Predicted
Cancer?	No	Yes
No	172	0
Yes	0	12

Contributing factors

Factor	Splits	Likelihood ratio χ^2	Effect
A10	90	4451.14056	0.2291
A8	83	3638.39958	0.1873

A3	66	3333.68573	0.1716
A7	58	1332.45026	0.0686
A12	40	1287.64646	0.0663
A4	40	1126.15477	0.0580
A11	25	981.501934	0.0505
A9	38	934.070021	0.0481
A15	59	913.067745	0.0470
A2	16	371.256757	0.0191
A1	33	341.492538	0.0176
A14	18	280.955495	0.0145
A13	8	172.026042	0.0089
A5	12	162.902991	0.0084
A6	14	101.888286	0.0052

Supplementary Table 4 Boosted tree model used for the prediction of colorectal adenomas

Overall statistics

	Measure	Training
Entropy	<i>R</i> -square	0.9269
Generalized	<i>R</i> -square	0.9630
Mean-Log	p	0.0481
Root-mean-squared-error		0.0571
Mean absolute deviation		0.0464
Misclassification rate		0.0000
<i>n</i>		184

Prediction

	Actual	Predicted
Adenoma?	No	Yes
No	116	0
Yes	0	68

Contributing factors

Factor	Splits	Likelihood ratio χ^2	Effect
A4	83	12685.8152	0.2008
A10	51	7240.88451	0.1146

A15	33	5916.23654	0.0936
A14	57	4786.45181	0.0758
A8	39	4561.1423	0.0722
A2	31	4018.13218	0.0636
A6	37	3504.59232	0.0555
A12	53	3477.97376	0.0551
A9	44	3461.08143	0.0548
A3	41	3027.65046	0.0479
A5	20	3018.25464	0.0478
A11	26	2924.4738	0.0463
A7	37	2336.48038	0.0370
A1	43	1802.11573	0.0285
A13	5	414.727285	0.0066

Supplementary Table 5 Boosted tree model used for the prediction of hyperplastic polyps

Overall statistics

	Measure	Training
	Entropy <i>R</i> -square	0.9947
	Generalized <i>R</i> -square	0.9962
	Mean-Log p	0.0017
	Root-mean-squared-error	0.0032
	Mean absolute deviation	0.0017
	Misclassification rate	0.0000
	<i>n</i>	184

Prediction

	Actual	Predicted
Hyperplastic polyp?	No	Yes
No	166	0
Yes	0	18

Contributing factors

Factor	Splits	Likelihood ratio χ^2	Effect
A7	96	7856.68623	0.2886
A1	89	6542.82426	0.2404

A4	77	5270.47961	0.1936
A8	61	1531.47147	0.0563
A2	40	1296.19397	0.0476
A6	48	1107.5793	0.0407
A14	47	796.719627	0.0293
A3	43	747.048464	0.0274
A10	27	609.962413	0.0224
A11	17	405.142335	0.0149
A12	17	358.159041	0.0132
A9	11	349.9098	0.0129
A5	17	147.393459	0.0054
A13	7	111.177224	0.0041
A15	3	89.5398425	0.0033

Supplementary Table 6 Boosted tree model used for the prediction of rectal neuroendocrine tumors

Overall statistics

	Measure	Training
Entropy	<i>R</i> -square	0.9985
Generalized	<i>R</i> -square	0.9986
Mean-Log	p	0.0001
Root-mean-squared-error		0.0003
Mean absolute deviation		0.0001
Misclassification rate		0.0000
<i>n</i>		184

Prediction

	Actual	Predicted	
Neuroendocrine tumor?		No	Yes
No		182	0
Yes		0	2

Contributing factors

Factor	Splits	Likelihood ratio χ^2	Effect
A11	166	3456.9728	0.7714
A12	33	296.727015	0.0662

A8	181	184.617998	0.0412
A5	11	145.214703	0.0324
A4	11	110.085833	0.0246
A3	9	81.5429748	0.0182
A7	5	80.5120268	0.0180
A1	14	66.1919133	0.0148
A2	161	27.7087675	0.0062
A14	2	21.9051656	0.0049
A9	3	9.76049583	0.0022
A15	4	0.21236603	0.0000
A6	0	0	0.0000
A10	0	0	0.0000
A13	0	0	0.0000

Supplementary Table 7 Statistical analyses of scattered light intensities among the patient groups

Raman shift	Pvalue (95% confidence interval)		
	A3	A8	A10
Groups			
Cancer <i>vs</i> adenoma	0.9998 (-18.9863, 25.26467)	0.7159 (-30.0877, 13.98800)	0.9685 (-23.1410, 39.04500)
Cancer <i>vs</i> HP	1.0000 (-26.55550, 31.25833)	0.9469 (-30.3637, 16.92267)	0.9792 (-29.6443, 47.52100)
Cancer <i>vs</i> other diseases and/or no specific findings	0.9993 (-19.7870, 23.88933)	0.9451 (-26.0770, 15.76133)	0.9793 (-23.0340, 41.06800)
Cancer <i>vs</i> NET	0.9911 (n/a)	0.9988 (n/a)	0.9988 (n/a)
Adenoma <i>vs</i> HP	0.9998 (-20.1713, 18.02033)	0.9996 (-16.0920, 16.58933)	0.9995 (-26.5083, 24.42967)
Adenoma <i>vs</i> other diseases and/or no specific findings	1.0000 (-12.3410, 11.40100)	0.9500 (-7.8010, 13.32300)	1.0000 (-17.1950, 14.74800)
Adenoma <i>vs</i> NET	0.9778 (n/a)	0.7420 (n/a)	0.9863 (n/a)

HP <i>vs</i> other diseases and/or no specific findings	1.0000 (-19.6537, 19.72000)	0.9951 (-14.8940, 18.19600)	1.0000 (-23.5067, 26.21267)
HP <i>vs</i> NET	0.9798 (n/a)	0.6769 (n/a)	0.9922 (n/a)
Other diseases and/or no specific findings <i>vs</i> NET	0.9809 (n/a)	0.8339 (n/a)	0.9920 (n/a)

HP: Hyperplastic polyp; NET: Neuroendocrine tumor; n/a: Not applicable.