

Table 1: Cohort and Group Clinicopathological Data

Characteristic		All Patients (n = 812)	No Early Recurrence (n = 576)	Early Recurrence (n =236)	P§
Gender	Male	687 (84.6)	487 (84.5)	200 (84.7)	0.944
	Female	125 (15.4)	89 (15.5)	36 (15.3)	
Age		64 (28-83)	63.9 (28-81)	64.1 (38-83)	0.855¶
Site	Oesophagus	361 (44.5)	250 (43.4)	111 (47)	0.352
	GOJ	451 (55.5)	326 (56.6)	125 (53)	
TRG	1-2	145 (17.9)	125 (21.7)	20 (8.5)	<0.001*
	3-5	667 (82.1)	451 (78.3)	216 (91.5)	
ypT	0	33 (2.6)	28 (4.9)	5 (2.1)	<0.001*
	1	96 (11.8)	87 (15.1)	9 (3.8)	
	2	141 (17.4)	125 (21.7)	16 (6.8)	
	3	495 (61.0)	320 (55.6)	175 (74.2)	
	4	47 (5.8)	16 (2.8)	31 (13.1)	
Positive LN (n)		1 (0-41)	0.5 (0-30)	4 (0-41)	<0.001*¶
ypN > 0		495 (61.0)	288 (50)	207 (87.7)	<0.001*
Total LN (n)		24 (0-75)	24 (0-75)	23 (6-61)	0.805¶
Total LN > 15		688 (84.7)	481 (83.5)	207 (87.7)	0.134
Lymphovascular Invasion		372 (45.8)	202 (35.1)	170 (72)	<0.001*
R1 Resection		236 (29.1)	118 (20.5)	113 (47.9)	<0.001*
Grade	Well	63 (7.8)	55 (9.5)	8 (3.4)	<0.001*
	Moderate	300 (36.9)	233 (40.5)	67 (28.4)	
	Poor/Anaplastic	449 (55.3)	288 (50)	161 (68.2)	
Neoadjuvant Treatment	NACT	657 (80.9)	476 (82.6)	181 (76.7)	0.061
	NACRT	155 (19.1)	100 (17.4)	55 (23.3)	

Data presented as absolute number (%) and median (range), *p<0.05. § χ^2 test, except ¶ Mann–Whitney U test, TRG – Tumour Regression Grade, GOJ – Gastro-oesophageal junction

Table 2: Model Discrimination AUC (95% CI)

Model	Apparent	Internal Validation	Internal-External Validation
Elastic Net	0.805 (0.772-0.838)	0.791 (0.757-0.826)	0.798 (0.713-0.883)
Random Forest	0.980 (0.972-0.987)	0.801 (0.769-0.834)	0.805 (0.721-0.889)
XG Boost	0.849 (0.822-0.877)	0.804 (0.772-0.836)	0.800 (0.716-0.883)
Ensemble	0.902 (0.881-0.992)	0.805 (0.790-0.819)	0.804 (0.721-0.887)

Table 3: Variable Importance

Predictor Variable	Elastic Net (%)	Random Forest (%)	XG Boost (%)	Final Model (%)
Gender	0.0	1.1	1.2	0.8
Age	0.3	18.2	10.2	9.6
Site of Tumour	9.4	2.6	4.8	5.6
Response to Neoadjuvant	0.0	0.0	0.0	0.0
ypT	11.2	9.2	7.4	9.2
Number of Positive LN	3.6	30.8	40.9	25.7
Total LN Examined	0.4	16.7	7.0	8.0
Lymphovascular Invasion	26.8	10.5	13.6	16.9
Completeness of Resection (R0/R1)	15.9	5.2	6.1	8.9
Grade	7.0	3.2	2.1	4.0
Neoadjuvant Treatment (CT/CRT)	25.4	2.5	6.8	11.4
CT – Chemotherapy, CRT - Chemoradiotherapy				

Table 4: Patient Examples using final model

Predictor Variable	Low Risk (%)	Medium Risk (%)	High Risk (%)
Baseline Prediction	27.4%	27.4%	27.4%
Gender	-0.1	-0.4	-2.0
Age	-0.8	-0.1	+4.3
Site of Tumour	-1.4	+9.8	+8.1
Response to Neoadjuvant Treatment	-0.5	+0.4	+0.1
ypT	-6.5	+4.9	+3.2
Number of Positive Lymph Nodes	-10.0	-32.2	+9.7
Nodes Examined	-1.2	-1.9	-3.1
Lymphovascular Invasion	-7.1	+21.1	+14.7
Completeness of Resection (R0/R1)	-2.3	-6.0	+6.1
Grade	-3.0	-6.9	+3.6
Neoadjuvant Treatment	+5.8	+22.2	-3.9
Final Prediction	0.3%	54.0%	68.2%

This table illustrates the percentage contribution of each variable in each example patient. This is represented in absolute percentage change from the mean predicted value of 27.4%. A calculator for this is packaged with the online model