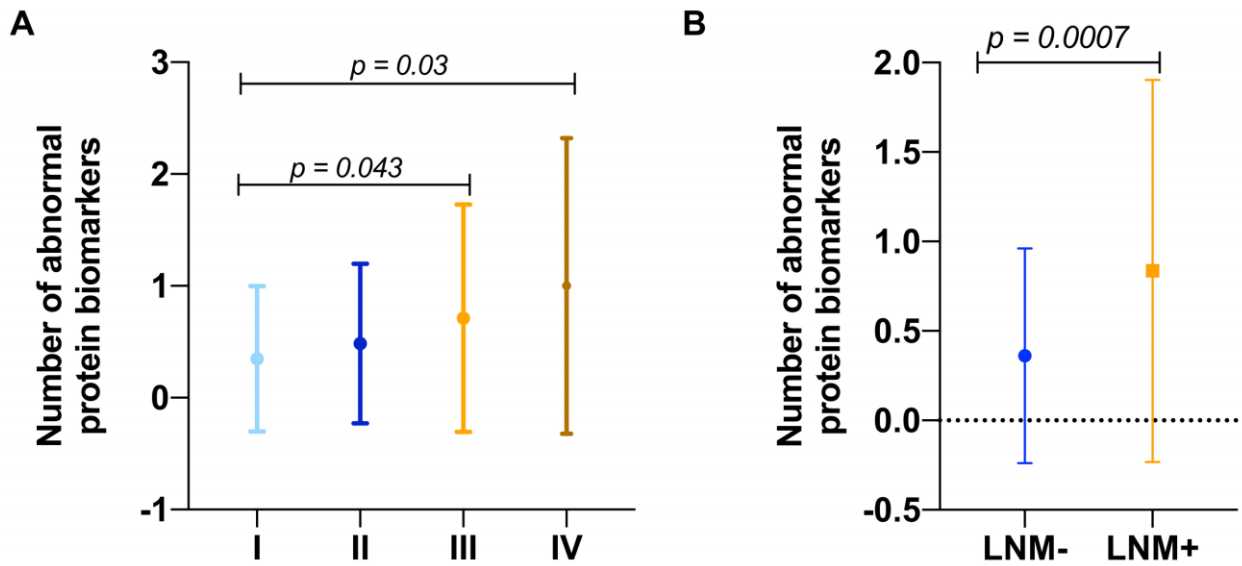


Supplementary Figure 1 The Kaplan-Meier survival analysis of esophageal cancer patients grouped by lymph node metastasis status. A: The disease-free survival of the whole cohorts; B: The disease-free survival of training cohorts; C: The disease-free survival of validation cohorts; D: The overall survival of the whole cohorts; E: The overall survival of training cohorts; F: The overall survival of validation cohorts.



Supplementary Figure 2 The relationship between number of abnormal protein biomarkers with tumor burden. A: The number of aberrant protein biomarkers in different stages; B: Comparison of number of aberrant protein biomarkers between the lymph node negative and positive groups.

Supplementary Table 1 Clinicopathological characteristics of the whole cohort, circulating tumor DNA cohort, and non-circulating tumor DNA cohort

Variables	Whole cohort (n = 206)	ctDNA cohort (n = 57)	Non-ctDNA cohort (n = 149)	P value
Age				0.391
mean (SD)	64.1 (7.5)	64.8 (7.4)	63.8 (7.5)	
Gender				0.356
Female	45	10	35	
Male	161	47	114	
Body mass index				0.713
mean (SD)	23.1 (2.7)	23.1 (2.6)	23.1 (2.7)	
Smoking and				0.532

Drinking				
None	86	22	64	
Smoking alone	40	9	31	
Drinking alone	13	3	10	
Both	67	23	44	
Tumor grade				0.213
G1	16	5	11	
G2	150	37	113	
G3	40	15	25	
T stage				0.654
T1	47	14	33	
T2	58	19	39	
T3	91	22	69	
T4	10	2	8	
N stage				0.520
N0	127	31	96	
N1	46	15	31	
N2	19	7	12	
N3	14	4	10	

Supplementary Table 2 Cross tables of lymph node metastasis status reported by tissue pathological and computed tomography

Lymph region	node Group	Computed tomography	Pathological		Total	Kappa	<i>P</i> value
			Positive	Negative			
Overall	Whole group	Positive	32	26	58	0.232	0.001
		Negative	44	104	148		

		Total	76	130	206		
	Training group	Positive	23	20	43	0.212	0.010
		Negative	31	70	101		
		Total	54	90	144		
	Validation group	Positive	9	6	15	0.279	0.023
		Negative	13	34	47		
		Total	22	40	62		
Thoracic	Whole group	Positive	23	22	45	0.347	< 0.001
		Negative	25	136	161		
		Total	48	158	206		
	Training group	Positive	19	16	35	0.371	< 0.001
		Negative	18	91	109		
		Total	37	107	144		
	Validation group	Positive	4	6	10	0.255	0.044
		Negative	7	45	52		
		Total	11	51	62		
Abdominal	Whole group	Positive	13	8	21	0.296	< 0.001
		Negative	32	153	185		
		Total	45	161	206		
	Training group	Positive	8	7	15	0.252	0.001

	Negative	22	107	129		
	Total	30	114	144		
Validation group	Positive	5	1	6	0.392	<
	Negative	10	46	56		
	Total	15	47	62		0.001

Supplementary Table 3 Univariate and multivariate logistic regression analysis of the training dataset

Variables	Variables type	Univariate analysis				Multivariate analysis			
		P value	OR	95%CI lower	95%CI upper	P value	OR	95%CI lower	95% CI upper
Age	numerical	0.55	1.01	0.97	1.05				
Gender	categorical	0.68	1.16	0.58	2.31				
Body mass index	numerical	0.10	0.91	0.82	1.02				
Alcohol unit	numerical	0.13	1.00	1.00	1.00				
Drinking	categorical	0.68	1.13	0.64	2.01				
Smoking index	numerical	0.65	1.00	0.99	1.01				
Smoking	categorical	0.89	1.04	0.59	1.83				
CA199 status	categorical	0.01	14.29	1.75	116.58	-	-	-	-
CA199 level	numerical	0.052	1.02	1.00	1.04				
CEA status	categorical	0.06	2.32	0.96	5.59				
CEA level	numerical	0.25	1.04	0.97	1.11				
CYFRA21-1 level	numerical	0.003	1.65	1.19	2.29	-	-	-	-
CYFRA21-1 status	categorical	0.03	2.42	1.10	5.32				
SCC level	numerical	0.02	1.61	1.10	2.36	-	-	-	-
SCC status	categorical	0.03	2.21	1.06	4.62				

Number of abnormal protein biomarkers	numerical	< 0.001	2.05	1.40	3.00	0.017	1.76	1.11	2.80
CT-reported nodal status	categorical	< 0.001	3.28	1.74	6.21	0.028	2.35	1.10	5.03
Tumor grade	categorical	0.43	1.25	0.72	2.18				
Tumor location	categorical	< 0.001	0.31	0.18	0.54	0.002	0.35	0.18	0.67
Tumor length	numerical	< 0.001	1.67	1.35	2.06	0.017	1.37	1.06	1.77
Depth of tumor invasion	categorical	< 0.001	3.37	2.20	5.15	< 0.001	2.72	1.71	4.32

Supplementary Table 4 Variables selected by different feature selection strategies

Variables	Set1-traditional univariate multivariate logistic regression analysis	Set2- and literature	Set3- combine	Set4- removing features with low variance	Set5- univariate feature selection	Set6- random forest	Set7- select from model	Set8- LASSO	Set9-RFECV
Tumor length	Y	Y	Y	N	Y	Y	Y	Y	Y
CT-reported nodal status	Y	Y	Y	Y	N	N	Y	Y	N
Tumor location	Y	Y	Y	Y	N	N	Y	Y	N
Depth of tumor invasion	Y	Y	Y	N	N	Y	Y	Y	N
Number of abnormal proteins	Y	N	Y	N	N	N	Y	N	N
Tumor grade	N	Y	Y	N	N	N	N	N	Y
CA199	N	N	N	Y	Y	Y	N	N	Y
Gender	N	N	N	Y	Y	N	N	N	N
SCC	N	N	N	N	Y	N	Y	N	Y
Drinking	N	N	N	N	Y	N	N	N	Y
CYFRA21-1	N	N	N	N	N	Y	Y	N	Y

BMI	N	N	N	N	N	N	N	N	Y
CEA	N	N	N	N	N	N	N	N	Y
Number of variables in the model	5	5	6	4	5	4	7	4	8

Supplementary Table 5 Univariate and multivariate logistic analysis of the circulating tumor DNA cohort

Variables	Variables type	Univariate analysis				Multivariate analysis			
		P value	OR	95%CI		P value	OR	95%CI	
				lower	upper			lower	upper
Age	Numerical	0.355	0.966	0.898	1.039				
Gender	Categorical	0.759	0.808	0.206	3.167				
BMI	Numerical	0.022	0.759	0.599	0.961	-	-	-	-
Smoking index	Numerical	0.447	1.008	0.987	1.029				
Smoking history	Categorical	0.646	1.278	0.448	3.652				
Drinking history	Categorical	0.03	3.36	1.128	10.008	-	-	-	-
Alcohol unit	Numerical	0.127	1	1	1.001				
CT-reported nodal status	Categorical	0.002	7.091	2.066	24.343	0.02	5.411	1.299	22.543

Tumor location	Categorical	0.007	0.158	0.042	0.599	0.027	0.186	0.042	0.828
Tumor length	Numerical	0.045	1.488	1.009	2.195	-	-	-	-
Tumor Grade	Numerical	0.453	1.423	0.567	3.571				
Number of abnormal protein biomarkers	Numerical	0.687	1.153	0.576	2.307				
Depth of tumor invasion	Categorical	0.997	1751319365	0	NR				
SCC level	Numerical	0.247	1.389	0.796	2.424				
CA199 level	Numerical	0.126	1.044	0.988	1.103				
CEA level	Numerical	0.549	1.097	0.811	1.482				
CYFR level	Numerical	0.553	1.181	0.681	2.047				
CYFRA21-1	Categorical	0.664	0.58	0.05	6.784				
SCC	Categorical	0.705	1.263	0.377	4.228				
CEA	Categorical	0.82	1.217	0.224	6.615				
CA199	Categorical	1	2003188805	0	NR				
Pre-surgical cfDNA concentration	Numerical	0.472	0.994	0.977	1.011				
ctDNA mutation counts	Numerical	0.945	0.991	0.772	1.272				

ctDNA mean VAF			Numerical	0.047	2.874	1.015	8.133	-	-	-	-
ctDNA TP53 mutation status			Categorical	0.18	2.109	0.709	6.279				
ctDNA TP53 mutation counts			Numerical	0.148	1.832	0.806	4.161				
ctDNA TP53 mean VAF			Numerical	0.009	2.943	1.303	6.647	0.040	2.41	1.042	5.573
ctDNA PIK3CA mutation status			Categorical	0.115	2.514	0.799	7.912				
ctDNA PIK3CA mutation counts			Numerical	0.329	1.454	0.685	3.087				
ctDNA PIK3CA mean VAF			Numerical	0.56	0.638	0.141	2.889				
cfDNA PTCH1 mutation status			Categorical	0.945	0.963	0.323	2.871				
cfDNA PTCH1 mutation counts			Numerical	0.869	0.934	0.418	2.09				
ctDNA PTCH1 mean VAF			Numerical	0.608	1.25	0.533	2.936				

Supplementary Table 6 Reported models for lymph node metastasis prediction in patients with esophageal cancer

Year	Tumor stage	Subjects number	Data source	Factors in models	Models in the article	Type of best-performing models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	PMID
2023	0-IV	3352	Patients' baseline characteristics, clinical information and tumor's histopathologic features	The pathology status of chest paraesophageal nodes, tumor depth, tumor location, side	9	Machine-learning-LightGBM	0.748	80.50%	44.60%	86.20%	/	90.70%	36864444
2020	Tis-T4	733	Patients' baseline characteristics, clinical information and tumor's histopathologic features	Tumor length, invasion depth, a past habit of alcohol use, histologic grade, lymph-vessel invasion, and preoperative CT results	2	Machine-learning-artificial neural network	91.50%	90.72%	87.06%	91.20%	56.49%	98.17%	33273861
2021	T1-T3	3186	Clinical features	Histology, clinical tumor stage, tumor size, grade, and presence of lymphovascular invasion	1	Nomogram	0.77	/	/	/	/	/	31274650
2010	0-III	273	Endoscopic features	Endoscopically determined clinical tumor depth (cT), clinical nodal (cN) status, and clinical tumor length (cL)	1	Nomogram	0.777	/	/	/	/	/	20881767
2021	T1	223 (training) + 80 (validation)	Physical examination, laboratory analysis, imaging examination and	Tumor differentiation, depth of tumor invasion and tumor length	1	Nomogram	0.830 (95% CI 0.763-0.902)	/	82%	76%	/	/	33879102

endoscopy

2018	T1	281 (training) + 198 (validation)	Clinical characteristics	Depth of tumour invasion, grade of differentiation, tumour length, lymphovascular invasion	1	Nomogram	0.801	/	47.20%	94.70%	/	/	29863776
2022	I-III	1097	Clinical features	/	36	Machine-learning-naive Bayes	0.752 (0.674- 0.829)	/	78.70%	63.80%	82.20%	58.50%	36158684
2019	T1-T3	366 (training) + 173 (validation)	Clinicopathological variables	Tumor length, tumor grade, invasion depth, angiolymphatic	1	Nomogram	0.80 (95% CI, 0.74- 0.86)	/	75.50%	74.80%	/	/	31029876
2017	T1	240	Pathological data	Tumor grade, T1 stage, tumor location and tumor length	1	Nomogram	0.762 (95%CI, 0.686- 0.838)	75.00%	67.00%	77.00%	36.00%	92.00%	29233596
2016	T1	210 EAC	Pathologic factors	Tumor length, grade, angiolymphatic invasion, and submucosal invasion	1	Multivariable logistic model	0.801 (95%CI, 0.726- 0.876)	/	/	/	/	/	26515637
2016	/	66 (training) + 65 (validation)	Computed tomography	/	1	Machine-learning-support vector machine	0.887	/	/	/	/	/	27879527
2022	/	530	Radiomic features of CT	Two-dimensional and three-dimensional size-based measurements	1	Machine-learning-support vector machine	0.841- 0.891	84.2- 94.7%	65.7-83.3%	84.4-96.7%	/	/	36018524

2018		197	Radiomic features of CT	Radiomic features, position	1	Nomogram	0.771 (95% CI: 0.632-0.910)	76.70%	61.20%	/	/	/	29727831
2022	/	380	Radiomic features and clinical features	/	9	Machine-learning-random forest	0.8,	76%	76%	76%	/	/	36147442
2018	/	154 (training) + 76 (validation)	1576 radiomics features extracted from arterial-phase CT images	Five radiomics features were reported LN status from CT	1	Nomogram	0.773	/	/	/	/	/	29922924
2014	I-II	143	18F-FDG parameters	The combination of clinical T classification and SUVmax	1	Receiver-operating-characteristic curve analysis	/	/	73.00%	81.50%	60.00%	89.70%	24700884
2020	/	411	Radiomic features of CT	Three radiomics features and CT-reported LN status	3	Nomogram	0.87%	/	/	/	/	/	32039021
2004	I - IV	36 (training) + 18 (validation)	Gene Expression of frozen tissue	44 genes	1	Machine-learning: neural networks combined with AdaBoosting	/	88.90%	/	/	/	/	15173069
2019	T1-T4	n = 56 + n=224	Gene-expression signature of mRNAs and clinical features	5-gene panel together with the lymphatic vessel invasion and venous invasion	3	Multivariate logistic regression model	0.74	75%	57%	92%	88%	69%	29240008