

Supplementary Table 1. Univariate logistic regression analysis of different variables in predicting positive lymph nodes in training set.

	OR	95%CI	P
Age			P<0.001
<45	1.97	1.74-2.23	
45-64	Reference	Reference	
>64	0.58	0.52-0.64	
Race			P<0.001
White	Reference	Reference	
Black	1.63	1.43-1.87	
Others	0.92	0.78-1.09	
Marital status			0.563
Married	Reference	Reference	
Unmarried	1.21	0.94-1.13	
Primary site			P<0.001
Central	1.13	0.91-1.41	
Inner	0.51	0.44-0.59	
Outer	Reference	Reference	
Tail	2.08	1.31-3.31	
Overlap	1.00	0.91-1.11	
Grade			P<0.001
I	0.45	0.39-0.52	
II	Reference	Reference	
III	2.17	1.97-2.40	
Histology			P<0.001
IDC	Reference	Reference	
ILC	0.86	0.71-1.03	

Others	0.74	0.64-0.85	
T stage			P<0.001
T1mi	0.19	0.10-0.36	
T1a	0.22	0.18-0.28	
T1b	0.36	0.36-0.41	
T1c	Reference	Reference	
ER status			P<0.001
ER+	Reference	Reference	
ER-	2.17	1.95-2.42	
PR status			P<0.001
PR+	Reference	Reference	
PR-	1.80	1.63-1.98	
HER2 status			P<0.001
HER2+	2.11	1.88-2.36	
HER2-	Reference	Reference	
Subtype			P<0.001
HR+/HER2-	Reference	Reference	
HR-/HER2+	3.07	2.53-3.62	
HR+/HER2+	2.06	1.80-2.36	
TNBC	2.06	1.80-2.36	

Abbreviations: IDC: invasive ductal carcinoma. ILC: invasive ductal carcinoma. ER: estrogen receptor. PR: progesterone receptors. HER2: human epidermal growth factor receptor 2. TNBC: Triple Negative Breast Cancer.

Notes: T1mic:0-1mm, T1a:2-5mm, T1b:6-10mm, T1c:11-20mm.