Supplementary Figure1.Predictive performance of the each model with clinical impact curves. A. Training set of RFC. B. Validation set of RFC. C. Training set of ANN. D. Validation set of ANN. E. Training set of DT. F. Validation set of DT. G. Training set of XGboost. H. Validation set of XGboost. I. Training set of SVM. J. Validation set of SVM. The blue curve (the high risk of the result) indicates the number of true positioning under threshold probability, and the black curve (the high risk) indicates the number of people classified into the positive (high risk) by the predictive model under each threshold probability.

Supplementary Table 1 LNM model performance in training and verification cohort

Training set	Sensitivity	95%CI-Lower	95%CI-Upper	Specificity	95%CI-Lower	95%CI-Upper	PPV	95%CI-Lower	95%CI-Upper	NPV	95%CI-Lower	95%CI-Upper
RFC	0.968	0.442	1.493	0.993	0.437	1.550	0.968	0.403	1.532	0.993	0.435	1.552
ANN	0.935	0.410	1.461	0.980	0.424	1.537	0.935	0.371	1.500	0.987	0.428	1.545
DT	0.871	0.346	1.396	0.967	0.411	1.524	0.871	0.306	1.435	0.974	0.415	1.532
XGboost	0.806	0.281	1.332	0.958	0.401	1.514	0.820	0.255	1.384	0.961	0.402	1.519
SVM	0.758	0.233	1.283	0.944	0.388	1.501	0.758	0.194	1.323	0.951	0.392	1.509
GLM	0.726	0.175	1.277	0.944	0.393	1.495	0.726	0.175	1.277	0.944	0.393	1.495
Radiologist	0.694	0.163	1.225	0.934	0.403	1.465	0.683	0.138	1.229	0.937	0.392	1.482

Testing set	Sensitivity	95%CI-Lower	95%CI-Upper	Specificity	95%CI-Lower	95%CI-Upper	PPV	95%CI-Lower	95%CI-Upper	NPV	95%CI-Lower	95%CI-Upper
RFC	0.966	0.440	1.491	0.961	0.405	1.518	0.848	0.284	1.413	0.992	0.433	1.551
ANN	0.897	0.371	1.422	0.946	0.389	1.502	0.788	0.223	1.352	0.976	0.417	1.535
DT	0.759	0.233	1.284	0.915	0.358	1.471	0.667	0.102	1.231	0.944	0.385	1.503
XGboost	0.621	0.095	1.146	0.899	0.343	1.456	0.581	0.016	1.145	0.913	0.355	1.472
SVM	0.552	0.026	1.077	0.876	0.319	1.433	0.500	-0.064	1.064	0.897	0.338	1.455
GLM	0.759	0.214	1.304	0.915	0.370	1.460	0.667	0.122	1.212	0.944	0.399	1.489
Radiologist	0.690	0.135	1.245	0.915	0.360	1.470	0.645	0.090	1.200	0.929	0.374	1.484

RFC.Random forest classifier. SVM.Support vector machine. DT. Decision Tree. ANN.Neural network. XGboost.Extreme gradient boosting. GLM. Generalized linear model.AUC.Area under curve. 95%CI.95% confidence interval. PPV.Positive prediction value, NPV.Negative prediction value, 95% CI.95% confidence interval.

Supplementary Table 2 Candidate variable screening is associated with LNM through the RFC algorithm.

Variables	Mean decrease accuracy	Mean decrease Gini
Age	3.07829046	0.01880905
Sex	-2.04E-35	4.21E-16
Site	-1.000500375	2.84E-15
Ulcer	-2.23E-35	1.94E-15
Gross type	1.000500375	6.33E-15
Tumor size	-5.43E-34	8.52E-16
Infiltration_depth	-5.18E-34	4.22E-16
Vascular invasion	-1.78E-34	7.54E-16
VTT	-4.98E-34	8.78E-16
TF	5.016719434	0.316246384
EV	2.560559841	0.009968113
Entropy	33.73892771	28.18862389
IG_all	13.18819116	5.290743606
IG_0	10.97114958	3.816035792
IG_45	12.26136629	3.559400984
IG_90	7.196012826	1.049981814
IV_all	4.943829797	0.05934451
IV_all_SD	6.506252398	0.140426466
IV_0	8.407867674	0.765617002
IV_45	8.169987886	1.227867206
IV_90	1.932685373	0.013700507
Haralick_all	21.57181865	17.89467075
Haralick_30	18.92859895	12.02480186
Haralick_45	6.067036791	0.954843876
Haralick_90	5.440000925	0.237201166
CSV	1.819598603	0.005958844

CP	1.951535811	9.14E-15

VVT. Vascular tumor thrombusTF. Total frequency;EV energy valueIgG all. Inverse gap full angle0. Inverse gapIggle 45. Inverse gap 45 degrees;Iggle 90 reverseAll. Inertia full angleAshland lands SD. Full angle SDIverle0 inertia valueIvchen 45 inertia value 45 degrees;Ivline 90 90 degrees inertia;All. Haralick full angleHaralick_ 0. Haralick 0 °; Haralick_ 45. Haralick 45; Haralick_ 90. Haralick 90; CSV cluster shadowCluster prominence.