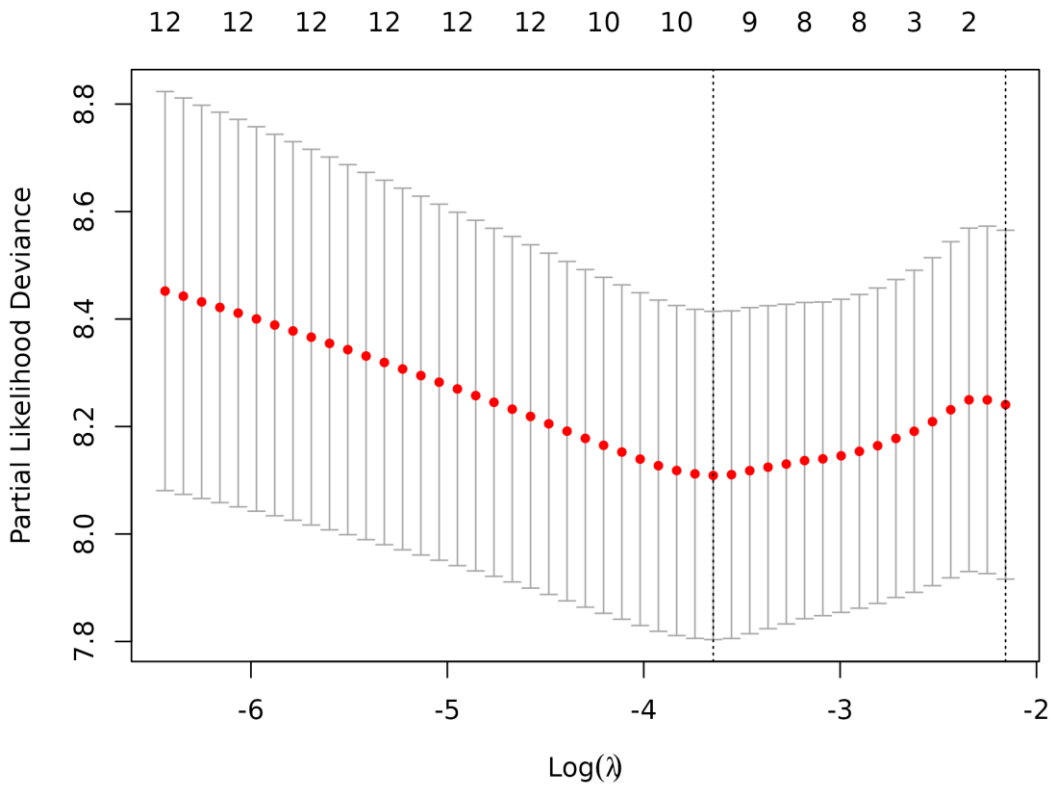
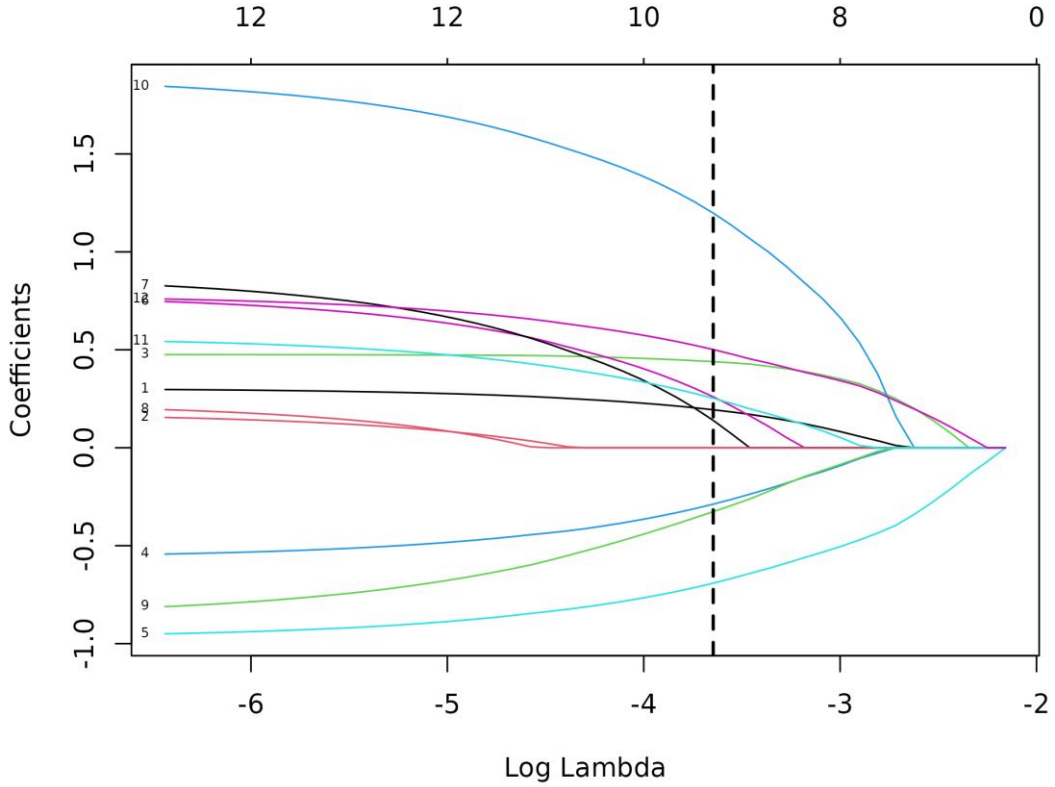
PART 1 Variable selection

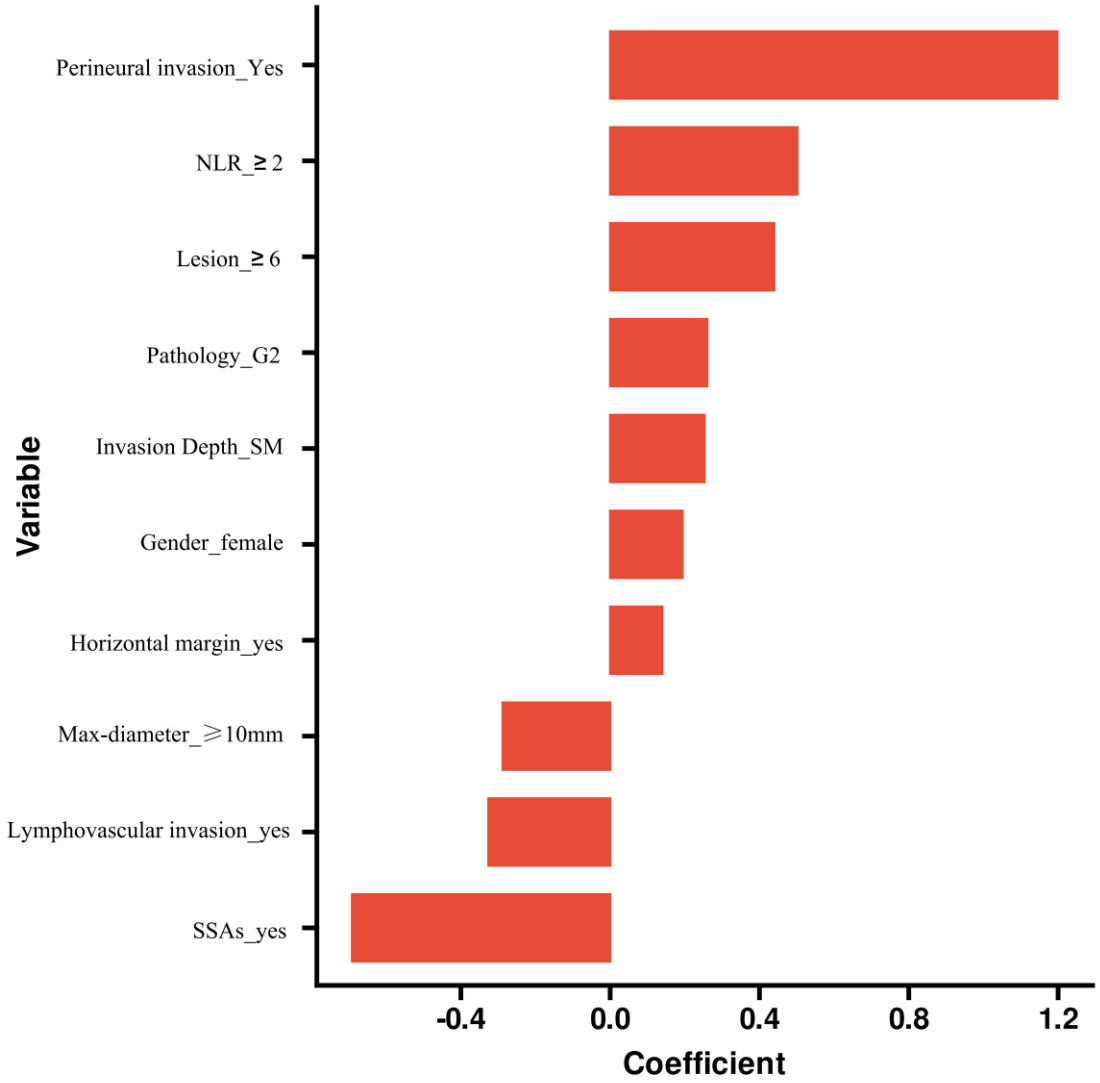
We employed Lasso regression (alpha=1) for variable selection, using 10-fold cross-validation and setting the number of lambda values to 100. The lambda value that minimized the mean squared error was used for variable screening. The following variables were selected by Lasso regression (when λ = 0.026091006791604) and can be used for subsequent modeling: Gender, Lesion, Max-diameter, SSAs, Pathology, Horizontal margin, Lymphovascular invasion, Perineural invasion, Invasion Depth, NLR.



Supplementary Figure 1 10-fold cross-validation was applied to select the most suitable feature using the Lasso regression model. (λ = 0.026091006791604).



Supplementary Figure 2 Plot of the Lasso regression coefficient profiles (λ = 0.026091006791604).

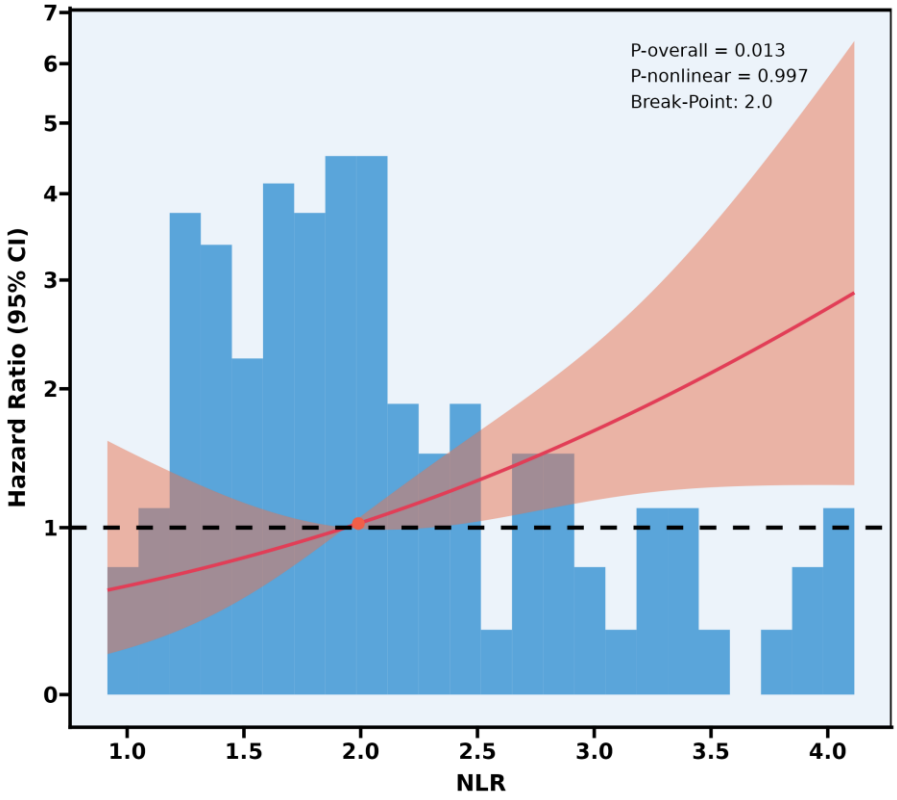


Supplementary Figure 3 Lasso-Selected Predictors and Corresponding Coefficients. Only variables with non-zero coefficients after Lasso selection are shown.

PART 2 Cutoff value of NLR

1.The overall analysis shows a statistically significant association with a p-value of 0.013, indicating that the relationship between NLR and the hazard ratio is meaningful.

2.In studying the relationship between NLR (neutrophil-to-lymphocyte ratio) and disease progression, we employed a piecewise Cox regression model with a fixed cut-off point. Using 2 as the cut-off value, we found a significant threshold effect of NLR on disease progression. When NLR was less than 2, the increase in the risk of disease progression was not significant (HR = 1.28, 95% CI = 0.46 – 3.61, P = 0.636). However, when NLR ⩾ 2, the risk of disease progression increased significantly (HR = 1.73, 95% CI = 1.14 – 2.64, P = 0.011). Compared with the standard Cox regression model, the piecewise model demonstrated a better fit (LRT test P = 0.648). This indicates the existence of a critical value for NLR, beyond which its impact on disease progression becomes significantly stronger.



Supplementary Figure 4 Association between NLR and mortality with the RCS function. Model with 3 knots located at 10th, 50th and 90th percentiles. Y-axis represents the HR to present mortality for any value of NLR compared to individuals with reference value (50th percentile) of NLR.

Supplementary Table 1 Threshold efect analysis of NLR on mortality

	HR (95% CI)*	P-value
Fitting by standard Cox regression model	1.62 (1.17, 2.24)	0.003
Fitting by piecewise Cox regression model (break-point = 2)		
NLR < 2	1.28 (0.46, 3.61)	0.636
NLR ≥ 2	1.73 (1.14, 2.64)	0.011
Log likelihood ratio		**0.648**

*Adjusted for: gender, lesionnumber, diagroup, SSA, pathology, horizon, lymphvascular, nerve, depth