Supplementary

Appendix 1 Radiomics features selection and model

The genetic algorithm (GA) was applied to further refine the selected feature subset from the previous steps. As an optimization technique inspired by natural selection, GA iteratively evolves feature combinations to enhance model performance. In this study, we set the population size to 100, the number of generations (NUM_EPOCHS) to 30, the crossover probability (cxpb) to 0.7, and the mutation probability (mutpb) to 0.2. Tournament selection (tournsize =3) was used for individual selection, while mutation was implemented as uniform integer mutation. Each candidate feature subset's performance was evaluated using 10-fold cross-validation during the GA process. The final feature subset was determined based on the overall performance improvement achieved during the optimization process.

Table S1 Breast MRI scanning parameters for the patients

Hospital	Scanner	Sequence	TR (ms)	TE (ms)	FOV (mm^2)	Slice thickness (mm)	Interslice gap (mm)	Acquisition time (s)
Center 1	1.5-T MRI system (uMR 560, United Imaging)	T1WI + C	5.1	2.1	320×320	2.4	0.48	394
		T2WI	3,800	42.7	328×350	4	0.8	126
		DWI	3,800	78.4	350×200	4	0.8	103
Center 2	1.5-T MRI system (Philips, Achieva)	T1WI + C	7.0	3.4	340×340	3	0	568
		T2WI	4,121	120	340×340	3	0	62
		DWI	5,366.4	68.9	340×340	5.5	0.4	85

DWI, diffusion-weighted imaging; FOV, field of view; MRI, magnetic resonance imaging; T1WI, T1-weighted imaging; T1WI + C, contrast-enhanced T1-weighted imaging; T2WI, T2-weighted imaging; TE, echo time; TR, repetition time.

Table S2 Construction of five models

Model	Defined
TA model	TA radscore
PA model	PA radscore
TPA model	TPA radscore
Clinical model	Clinical factors
Combined model	Clinical factors + highest radscore (TPA radscore)

PA, peritumoral area; radscore, radiomics score; TA, tumor area; TPA, tumor-plus-peritumoral area.

https://dx.doi.org/10.21037/qims-2024-2685

Table S3 Interobserver agreement for evaluating MRI characteristics

MRI characteristics	ICC value	95% CI	p value
LD	0.980	0.970–0.980	<0.001
TIC	0.877	0.817–0.918	<0.001
MRI-ALN status	0.752	0.630–0.833	<0.001
PE	0.790	0.687–0.859	<0.001
Multifocal or multicentric	0.815	0.724–0.875	<0.001
Shape	0.821	0.733–0.880	<0.001
Margin	0.866	0.801–0.910	<0.001
Internal enhancement	0.705	0.560–0.802	<0.001
FGT	0.747	0.624–0.830	<0.001

ALN, axillary lymph node; CI, confidence interval; FGT, fibroglandular tissue; ICC, intraclass correlation coefficient; LD, largest diameter; MRI, magnetic resonance imaging; PE, Peritumoral edema; TIC, time-signal intensity curve.

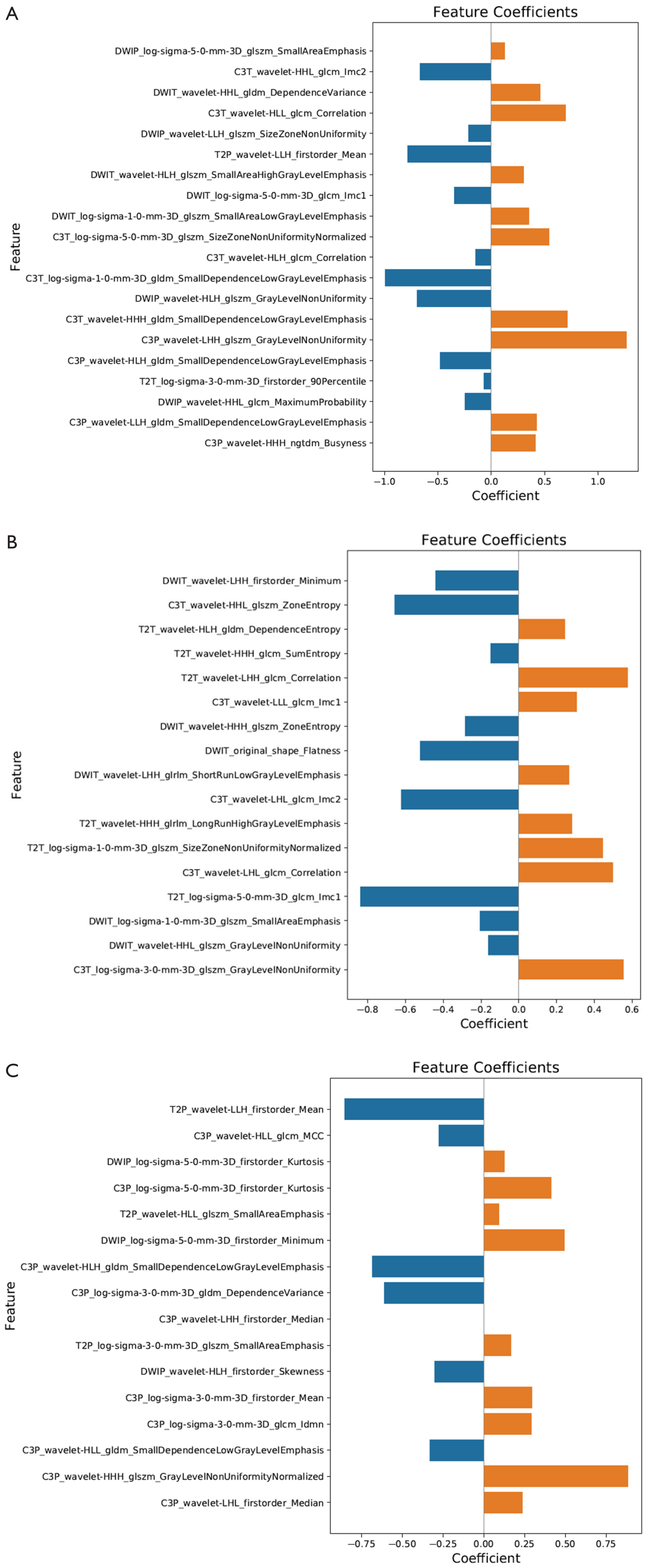

Figure S1 The bar chats of coefficients of TPA features (A), TA features (B), PA features (C), for distinguishing different LVI status. LVI, lymphovascular invasion; PA, peritumoral area; TA, tumor area; TPA, tumor-plus-peritumoral area.

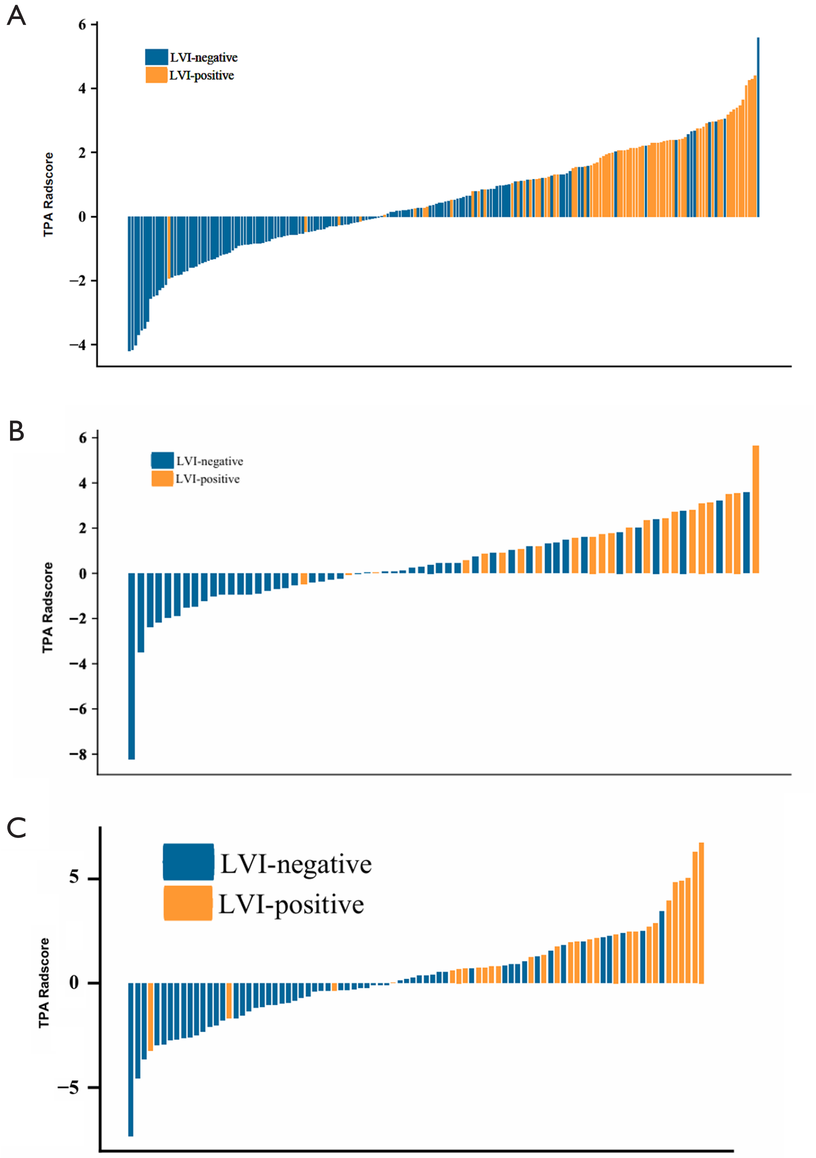

Figure S2 The TPA radscore for each patient in training, validation and test sets (A-C), respectively. LVI, lymphovascular invasion; radscore, radiomics score; TPA, tumor-plus-peritumoral area.

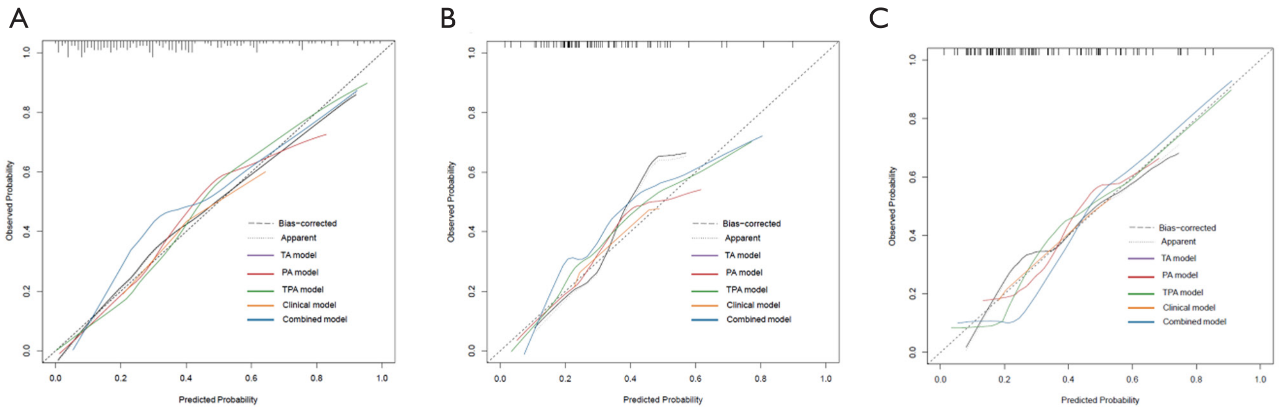

Figure S3 The calibration curves of the different models in training, validation and test sets. The X-axis represents different models predicted probability of LVI, the Y-axis an actual LVI status, and the diagonal dashed line indicates the ideal prediction by a perfect model (A-C). LVI, lymphovascular invasion; PA, peritumoral area; TA, tumor area; TPA, tumor-plus-peritumoral area.