Appendix 1: Deep learning feature definitions

Data preprocessing

With the tumor region clearly delineated, we extracted 3 consecutive axial slices that contained the largest amount of tumor tissue. These slices were then cropped to a size of 224 mm × 224 mm using a bounding box that encompassed the entirety of the tumor. This size corresponded to the input layer of the models used. The cropped images with 3 consecutive axial slices as image channels were used as the input for the convolutional neural network (CNN) model.

Convolutional neural network architecture

In our study, ResNet50 (25) was used for the extraction of representational deep learning features. This network was pretrained on ImageNet (47). This publicly released dataset contains a substantial number of object categories and manually annotated training images. The optimization hyperparameters were not tuned, which meant a broader generalization on the other datasets. The models are publicly assessable using Keras and TensorFlow open-source code (https://github.com/fchollet/deep-learning-models/releases/download/) under the MIT license. After preprocessing, 3 consecutive slices in computed tomography (CT) images with the maximum area of the tumor lesion were propagated in the network to generate deep learning features.

Removal of the last fully connected layer

For the pretrained models, the convolutional base is connected by a fully connected layer. We removed the last fully connected layer. A total of 2048 feature maps were obtained from the new output of this model.

Addition of a max pooling layer and feature extraction

With the use of a global pooling window, local data are concentrated, thus decreasing dimensionality. After Step 1.3, for models with more than 1 dimensional feature, we obtained feature maps with height and width dimensions consistent with the location invariance in the input layer. Following global pooling, the feature map vectors were transformed to their respective maximum raw values. The feature maps were transformed to numeric values, which were the representational deep learning features.

Appendix 2: Parameters of CT images and follow-up time

Follow-up time for overall survival

The overall survival (OS) is often regarded as the best endpoint of interest in survival analysis. For our study, the endpoint of follow-up was January 2022. The time from diagnosis to death or the end of follow-up was recorded as OS in our study. The median follow-up was 29.5 months, and the maximum was 124 months. In the first 2 years, the follow-up occurred every 2 or 3 months. Thereafter, follow-up occurred every 6 months. The follow-up process involved checking inpatient medical records, outpatient return records, and making phone calls to collect follow-up data.

Table S1 CT scanning equipment and scanning scheme

	Parameters	Scheme
Equipment	Equipment name	GE Discovery 750 HD CT
		Philips Brilliance iCT
		GE BrightSpeed CT
		Siemens Somatom Perspective CT
Scanned protocol	Width of collimator	32×0.6 mm or 64×0.625 mm
	Rotation time	0.5–0.8 s/r
	Tube voltage	120 kVp
	Tube current	290–650 mA
	Pitch	1.375:1/0.992:1
	Layer thickness/spacing	5.0 mm/5.0 mm
	Matrix	512×512
	Noise figure	10 HU
Scanned area	Location standard	Top of diaphragm to lower pole of both kidneys
Enhancement condition	Contrast agent	Iohexol (300 mgI/mL) or ioversol (320 mg/mL)
	Flow rate	2.5–3.5 mL/s
	Dose	1.5 mL/kg
	Acquisition time	Arterial phase: 30 s; venous phase: 60–70 s
Post-process	Reconstruction thickness	0.625 mm, 1.25 mm

CT, computed tomography; HU, Hounsfield unit.

https://dx.doi.org/10.21037/qims-23-577

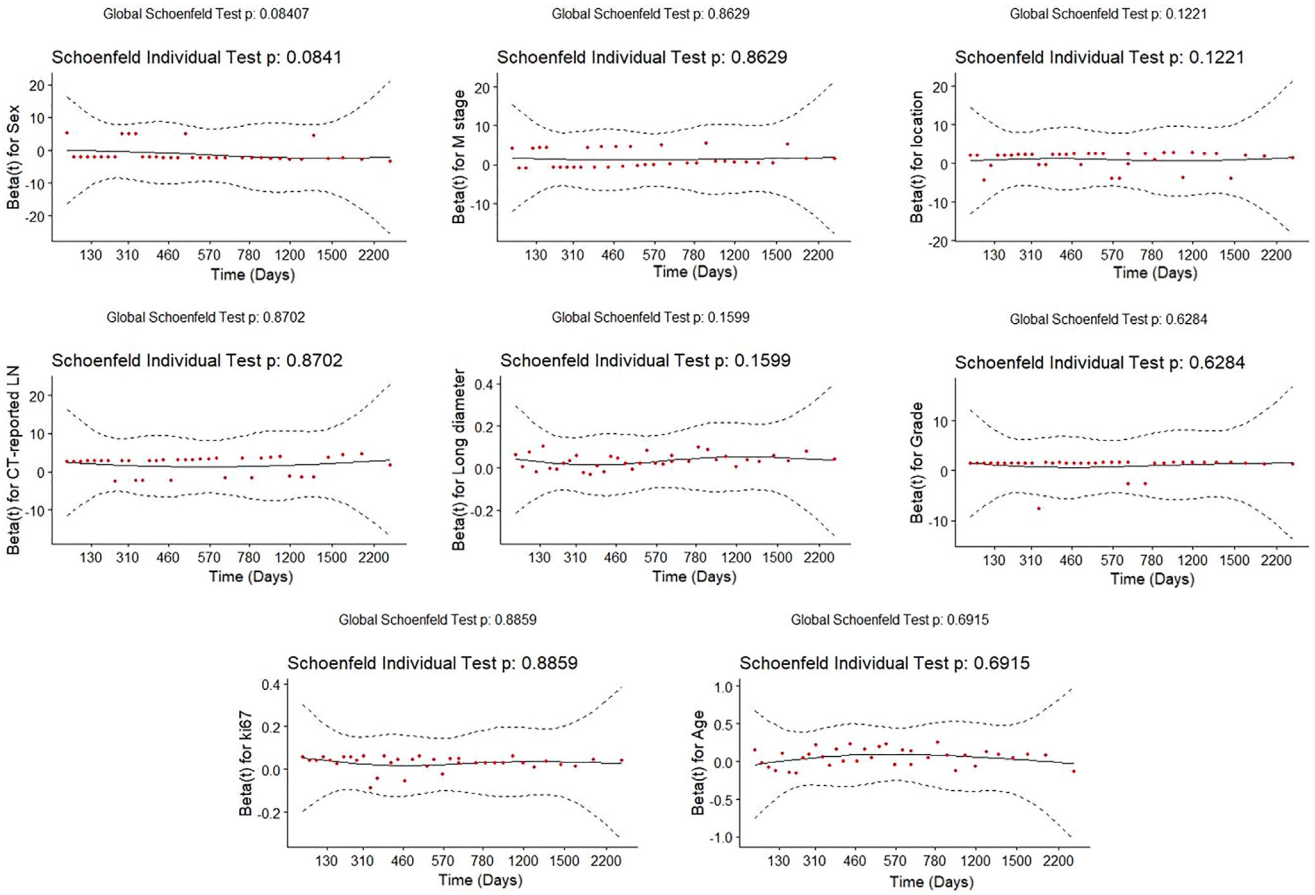

Figure S1 Validation of the proportional hazard assumption. For each clinical co-variable, Schoenfeld residuals test with chi-squared test was calculated (48). Factors with P>0.05 were considered eligible for Cox regression. For each of the covariables in the Cox model, the P value was not statistically significant, and the P value for the global test was also not statistically significant. Therefore, it was reasonable to use Cox regression for univariable and multivariable analysis.

https://dx.doi.org/10.21037/qims-23-577

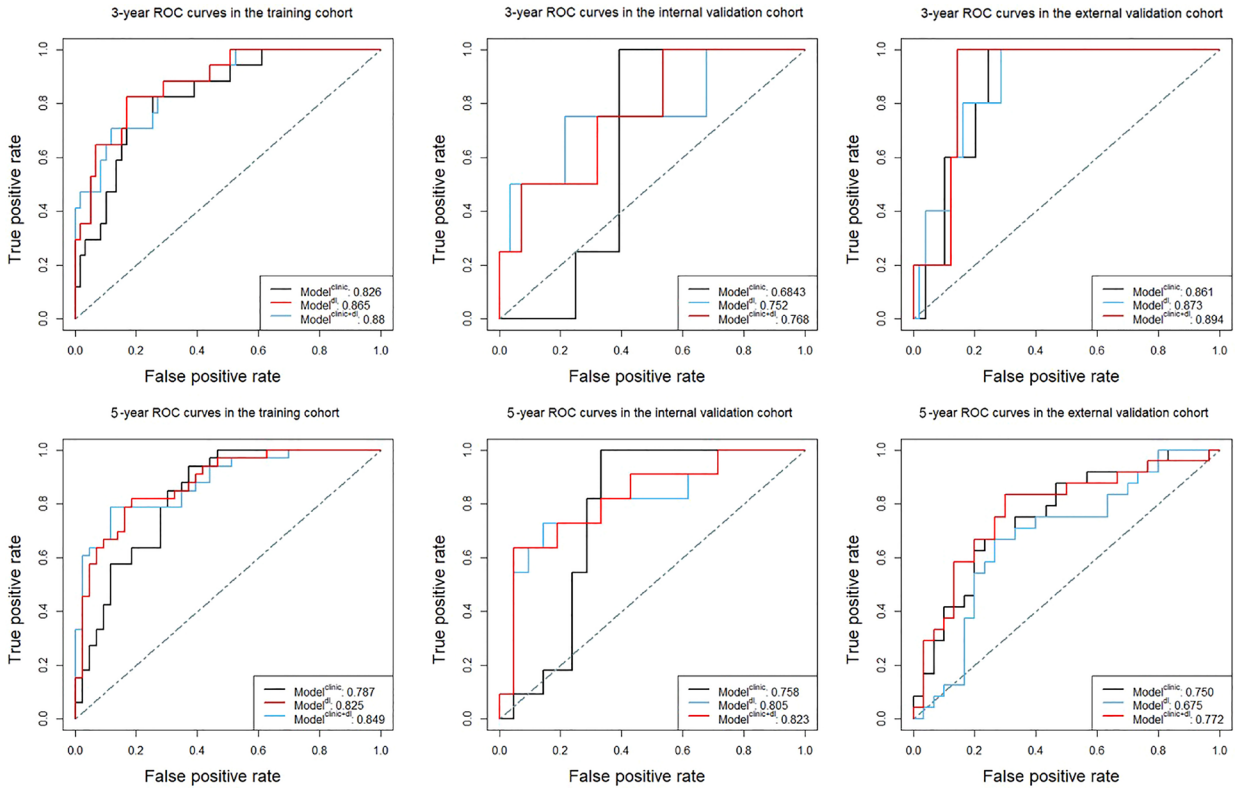

Figure S2 Time-dependent receiver operating characteristic curves for 3 and 5 years for each of the 3 models in the training, internal validation, and external validation cohorts. Receiver operating characteristic curves are shown for 3 cohorts.

References

47. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang ZH, Karpathy A, Khosla A, Bernstein MS, Berg AC, Li FF. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 2015;115:211-52.
48. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81:515-26.

https://dx.doi.org/10.21037/qims-23-577