SUPPLEMENTARY MATERIALS

Alternative Multivariate Modeling Approaches

The primary machine learning approach to developing the prognostic indices for OS was the LASSO-penalized Cox model, which is similar to prior work (1). The performance of LASSO-based models tend to perform better on independent datasets than regression models developed using standard stepwise feature selection techniques (2). Due to the substantial collinearity among the texture feature, we used a backwards elimination routine to reduce collinearity (all retained variables have R^2 <50%) between texture features before including the selected features in a LASSO penalized Cox model (termed Colin<50% + LASSO). There are alternative approaches to reducing collinearity among predictors, including those based on mutual information, partial least squares, principal components, and variable clustering (3-5). The primary advantage of the model development approach used in this study is the transparency and interpretability of the final model, in which a relatively small set of individual features was selected and each feature was given a clear weight in the prediction.

To investigate the impact of choice of model development approach on performance, several alternative feature selection approaches and model types were also considered using the derivation cohort. The two other feature selection approaches were 1) to select the best texture feature per feature class (histogram, GTSDM, NGTDM, and GLZSM) based on the smallest p-value from a univariate Cox model (termed Pick1) and 2) select all texture features with a p-value < 0.05 from the univariate analysis (termed Uni<0.05). The two other model types considered were 1) a penalized Cox model using the ridge regression (RR) penalty and 2) a Cox model using partial least squares (PLS) components as predictors (4), both of which may be more robust in the presence of substantial collinearity than the LASSO approach. As with the LASSO-penalized approach, the RR penalty parameter was selected to minimize the partial likelihood deviance calculated using leave-one-out cross-validation. In PLS, orthogonal components (linear combinations of features) are generated, similar to those generated during principal component analysis (PCA). However, unlike for PCA, PLS components are chosen to maximize their covariance with the outcome variable. In this application, up to 4 PLS components were selected for inclusion in a Cox model, where the number of components was selected to maximize the c-index calculated using 5-fold cross-validation.

In addition to the primary approach (Colin<50% + LASSO), six other combinations of these feature selection and model fitting approaches were constructed and used with the derivation cohort: RR alone (no feature selection), PLS alone, Pick1 + RR, Pick1 + PLS, Uni<0.05 + RR, and Uni<0.05 + PLS. The resulting models, apparent c-index (c-index of the derived model calculated on derivation cohort), bootstrap-adjusted c-indices are summarized in Supplementary Table 1. The apparent and bootstrap-adjusted performance of the primary model was 0.79 and 0.74, respectively. The corresponding measures of performance of the alternative models ranged from 0.76-0.78 and 0.72-0.74, respectively, so none had a notable performance advantage over the primary approach.

Supplementary Table 1. Multivariate models for OS using different methods based on the derivation dataset.

	Hazard Ratio*						
	Model 1:	**M2:**	**M3:**	**M4:**	**M5:**	**M6:**	**M7:**
Feature Selection:	**Colin<50%+**	***<None>***	***<None>***	**Pick1+**	**Pick1+**	**Uni<0.05+**	**Uni<0.05+**
Clinical Variables **Model Type:**	**LASSO**	**RR**	**PLS**	**RR**	**PLS**	**RR**	**PLS**
Age	1.4	1.2	1.5	1.3	1.5	1.3	1.5
Grade 3 (vs. grade 1-2)	1.7	1.3	0.9	1.5	0.9	1.4	0.9
Radiomics: Tumor Volume							
Tumor volume†	1.5	1.2	1.2	1.3	1.3	1.2	1.2
Radiomics: Histogram Variables							
Coefficient of variation		-	-				
Skewness	-	-	-			-	-
Kurtosis†	1.2	1.1	1.1				
Energy†		1.1	1.1			1.1	1.1
Entropy		0.9	0.9	0.9	0.9	0.9	0.9
Radiomics: GTSDM Variables							
Angular second moment/energy†		-	-			-	-
Contrast†		0.9	0.9	0.9	0.9	0.9	0.9
Correlation		-	-			-	-
Sum of squares variance		0.9	0.9			0.9	0.9
Inverse difference moment/ Homogeneity		-	-			-	-
Entropy		-	-			-	-
Autocorrelation		0.9	0.9			0.9	0.9
Dissimilarity		-	-			-	-
Radiomics: NGTDM Variables							
Coarseness†		0.9	0.9			0.9	0.9
Contrast†		-	-			-	-
Busyness†		1.1	1.1	1.2	1.3	1.1	1.1
Complexity	-	1.1	1.1				
Texture Strength†		-	-			-	-
Radiomics: GLZSM Variables							
Small zone size emphasis	-	-	1.1			1.1	1.2
Large zone size emphasis†		-	-			-	-
Low gray-level zone emphasis†		-	1.1				
High gray-level zone emphasis		-	-			-	0.9
Small zone / low gray emphasis†	1.2	-	1.1				
Small zone / high gray emphasis		-	-			-	-
Large zone / low gray emphasis†		1.1	1.1	1.1	1.2	1.1	1.1
Large zone / high gray emphasis†		0.9	0.9			0.9	0.9
Gray-level non-uniformity†		1.1	1.2			1.1	1.1
Zone size non-uniformity†	1.3	1.1	1.2			1.1	1.2
Zone size percentage		-	1.1			-	1.1
Model Performance							
Apparent C-index	0.79	0.78	0.77	0.76	0.76	0.78	0.78
.632 Bootstrap C-index	0.74	0.72	0.72	0.74	0.73	0.73	0.73
(95% CI)	(0.66, 0.80)	(0.64, 0.79)	(0.64, 0.79)	(0.66, 0.81)	(0.65, 0.80)	(0.66, 0.80)	(0.65, 0.81)
P-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Note: See the supplemental methods for descriptions and definitions of each model. Model 1 (Colin<50% + LASSO) is the primary approached used in the main text.

*Hazard ratio (HR) is per 1 SD-increase for continuous variables; HR > 1 indicates higher risk of death; HR values which round to 1.0 were replaced with "-" to indicate they were affectively "de-selected";

†Variable was log-transformed or cube-rooted before entry into the model to reduce right-skewness.

Supplementary Table 2. Univariate associations of clinical and radiomic features with overall survival based on the derivation data set.

Clinical Variables	C-index	HR*	(95% CI)	P-value†
Male sex	0.56	1.6	(0.8, 3.2)	0.17
Age	0.64	1.7	(1.2, 2.3)	**0.003**
Grade 3 (vs. grades 1-2)	0.60	2.4	(1.1, 5.1)	**0.022**
Extremity location	0.50	0.9	(0.5, 1.9)	0.86
Positive margins‡	0.54	0.7	(0.3, 1.6)	0.39
Pathology size > 5 cm	0.54	1.8	(0.6, 5.1)	0.27
Radiomics: Tumor Volume				
Tumor volume§	0.69	2.3	(1.5, 3.5)	**<0.001**
Radiomics: Histogram Features				
Coefficient of variation	0.65	1.1	(0.8, 1.6)	0.49
Skewness	0.59	1.5	(1.1, 2.0)	**0.015**
Kurtosis§	0.57	1.3	(1.0, 1.8)	0.083
Energy§	0.63	1.5	(1.1, 2.0)	**0.006**
Entropy	0.66	0.6	(0.5, 0.9)	**0.002**
Radiomics: GTSDM Features				
Angular second moment/energy§	0.61	1.5	(1.1, 2.0)	**0.013**
Contrast§	0.66	0.6	(0.4, 0.8)	**0.001**
Correlation	0.63	1.5	(1.1, 2.2)	**0.015**
Sum of squares variance	0.66	0.6	(0.4, 0.8)	**0.002**
Inverse difference moment/ homogeneity	0.63	1.6	(1.2, 2.1)	**0.003**
Entropy	0.63	0.7	(0.5, 0.9)	**0.005**
Autocorrelation	0.65	0.6	(0.4, 0.8)	**0.002**
Dissimilarity	0.65	0.5	(0.4, 0.8)	**0.002**
Radiomics: NGTDM Features				
Coarseness§	0.68	0.5	(0.4, 0.7)	**<0.001**
Contrast§	0.64	0.7	(0.5, 0.9)	**0.009**
Busyness§	0.70	2.1	(1.4, 3.1)	**<0.001**
Complexity	0.53	0.9	(0.7, 1.2)	0.53
Texture Strength§	0.64	0.6	(0.4, 0.8)	**0.003**
Radiomics: GLZSM Features				
Small zone size emphasis	0.61	0.7	(0.5, 1.0)	**0.034**
Large zone size emphasis§	0.65	1.7	(1.2, 2.3)	**0.001**
Low gray-level zone emphasis§	0.55	1.2	(0.9, 1.6)	0.27
High gray-level zone emphasis	0.65	0.6	(0.4, 0.9)	**0.005**
Small zone / low gray emphasis§	0.53	1.1	(0.8, 1.5)	0.58
Small zone / high gray emphasis	0.66	0.6	(0.4, 0.8)	**0.003**
Large zone / low gray emphasis§	0.68	1.9	(1.4, 2.6)	**<0.001**
Large zone / high gray emphasis§	0.62	1.4	(1.1, 2.0)	**0.022**
Gray-level non-uniformity§	0.68	2.0	(1.4, 2.9)	**<0.001**
Zone size non-uniformity§	0.66	1.9	(1.3, 2.7)	**<0.001**
Zone size percentage	0.64	0.6	(0.4, 0.9)	**0.007**

*Hazard ratio (HR) is per 1 SD-increase for continuous variables; HR > 1 indicates higher risk of death;
†P-values for Wald test of HR = 1 without adjustment for multiple comparisons;
‡Margins (n=9) were missing in some cases and excluded from the corresponding row of the univariate analysis;
§Variable was log-transformed or cube-rooted before entry into the model to reduce right-skewness.

Supplementary Table 3. Collinearity statistics of radiomics variables based on the derivation data set.

	All Variables		R^2 < 90%*		R^2 < 80%*		R^2 < 70%*		R^2 < 60%*		R^2 < 50%*	
Radiomics: Histogram Variables	**R^2**	**VIF**	**R^2**	**VIF**	**R^2**	**VIF**	**R^2**	**VIF**	**R^2**	**VIF**	**R^2**	**VIF**
Coefficient of variation	97.3%	37.1	84.7%	6.5								
Skewness	88.8%	8.9	79.2%	4.8	55.2%	2.2	54.3%	2.2	50.9%	2.0	40.0%	1.7
Kurtosis†	84.8%	6.6	75.9%	4.2	62.4%	2.7	61.5%	2.6	41.6%	1.7	35.7%	1.6
Energy†	99.5%	188.7	89.4%	9.5								
Entropy	99.4%	179.1										
Radiomics: GTSDM Variables												
Angular second moment/energy†	99.5%	213.1										
Contrast†	99.1%	114.7										
Correlation	97.0%	33.1	70.2%	3.4	60.3%	2.5	57.0%	2.3	55.4%	2.2		
Sum of squares variance	100.0%	71337.0										
Inverse difference moment/ homogeneity	98.3%	57.7										
Entropy	99.5%	205.1										
Autocorrelation	100.0%	69245.6										
Dissimilarity	99.8%	424.1										
Radiomics: NGTDM Variables												
Coarseness†	99.7%	325.2										
Contrast†	96.4%	28.0	86.4%	7.4	75.1%	4.0	68.7%	3.2				
Busyness†	99.7%	345.4										
Complexity	88.2%	8.5	56.3%	2.3	55.3%	2.2	54.3%	2.2	43.8%	1.8	29.4%	1.4
Texture Strength†	99.0%	97.2										
Radiomics: GLZSM Variables												
Small zone size emphasis	98.2%	56.5	80.5%	5.1	51.6%	2.1	48.4%	1.9	47.8%	1.9	45.4%	1.8
Large zone size emphasis†	99.9%	708.8										
Low gray-level zone emphasis†	99.3%	141.4										
High gray-level zone emphasis	99.8%	405.4										
Small zone / low gray emphasis†	99.1%	117.5	82.9%	5.8	47.0%	1.9	44.0%	1.8	43.9%	1.8	43.9%	1.8
Small zone / high gray emphasis	99.7%	295.2	87.8%	8.2								
Large zone / low gray emphasis†	99.5%	203.3										
Large zone / high gray emphasis†	99.6%	264.6	79.9%	5.0	77.0%	4.4						
Gray-level non-uniformity†	99.8%	516.4										
Zone size non-uniformity†	99.8%	452.3	63.1%	2.7	60.8%	2.6	59.1%	2.4	51.9%	2.1	44.9%	1.8
Zone size percentage	98.5%	67.2										
No. of variables	29		12		9		8		7		6	

*Variables shown are those that remained after applying backwards elimination until remaining variables met the R^2 criterion;

†Variable was log-transformed or cube-rooted before entry into the model to reduce right-skewness.

Supplementary Table 4. Discrimination performance of models in the derivation and validation cohorts.

	Center 1 OS ≤ 3Y (37 deaths)		Center 2 OS ≤ 3Y (16 deaths)		Center 2 OS (21 deaths)	
Model	**C-index***	**(95% CI)**	**C-index**	**(95% CI)**	**C-index**	**(95% CI)**
Model C: Age + grade	0.68	(0.60, 0.75)	0.70	(0.57, 0.82)	0.71	(0.61, 0.82)
Model R: Radiomics only	0.69	(0.61, 0.76)	0.68	(0.55, 0.80)	0.68	(0.56, 0.79)
Model C+R: Age + grade + radiomics	0.74	(0.66, 0.80)	0.77	(0.64, 0.88)	0.78	(0.66, 0.88)
Model R_A: Tumor volume only	0.70	(0.62, 0.76)	0.68	(0.55, 0.80)	0.69	(0.57, 0.80)
Model R_B: Texture features only	0.67	(0.58, 0.74)	0.65	(0.51, 0.78)	0.65	(0.52, 0.78)
Model C+R_A: Age + grade + tumor volume	0.75	(0.68, 0.82)	0.75	(0.63, 0.86)	0.77	(0.66, 0.87)
Model C+R_B: Age + grade + texture features	0.72	(0.64, 0.79)	0.76	(0.62, 0.88)	0.77	(0.65, 0.88)

C = Clinical only, R = Radiomics only, R_A = tumor volume only; R_B = Texture Features; C+R=Clinical + Radiomics; C+R_A = Clinical + Tumor Volume only; C+R_B = Clinical + Texture Features;

*Estimated using the .632 bootstrap.

REFERENCES

1. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
2. Ojeda FM, Müller C, Börnigen D, et al. Comparison of Cox Model Methods in A Low-dimensional Setting with Few Events. Genomics Proteomics Bioinformatics. 2016;14(4):235-43.
3. Leger S, Zwanenburg A, Pilz K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. Sci Rep. 2017;7(1):13206.
4. Bastien P, Bertrand F, Meyer N, Maumy-Bertrand M. Deviance residuals-based sparse PLS and sparse kernel PLS regression for censored data. Bioinformatics. 2015;31(3):397-404.
5. Bair E, Hastie T, Paul D, Tibshirani R. Prediction by supervised principal components. Journal of the American Statistical Association. 2006;101(473):119-37.