

Supplementary Table 1. Image acquisition characteristics

Cohort	Scanner	#Scans	KVP	In-plane resolution	Slice thickness	Contrast
LiTS cohort	N/A	131	N/A	0.768 (0.557,1)	<1 (51) 1 (38) 1-5 (33) 5 (9)	None (1) Venous (130)
mSCLC segmentation (n = 82 series volumes)	SIEMENS	73	120	0.820 (0.691-0.977)	1 (3)	None (5) Venous (66) Arterial (2)
	Canon Medical Systems	3	120	0.858 (0.782-0.858)	1 (1) 5 (2)	Venous
	GE Medical Systems	1	120	0.7813	1-5 (1)	Venous
	TOSHIBA	5	120	0.938 (0.743-0.938)	1 (1) 5 (4)	Arterial (1) Venous (4)
mSCLC classification (n = 227 series volumes)	SIEMENS	207	120 (205) 100 (2)	0.820 (0.677-0.977)	1 (105) 5 (102)	None (7) Venous (200)
	Canon Medical Systems	12	120	0.839 (0.782, 0.9)	1 (6) 5 (6)	Venous
	GE Medical Systems	2	120	0.7813, 0.7031	1-5 (2)	Arterial (1) Venous (1)
	TOSHIBA	6	120	0.782 (0.743, 0.782)	1 (3) 5 (3)	Arterial (1) Venous (5)

LiTS, liver tumor segmentation; mSCLC, metastatic small cell lung cancer; KVP, kilovolt peak; N/A, not available.

Supplementary Table 2. Cross-validation performance across weighted and unweighted models

Weighted	Optim	Pos class weight	Model	F1	AUC
No	-	-	Logistic regression	0.85 (0.81, 0.89)	0.55 (0.45, 0.67)
No	-	-	Random forest	0.86 (0.81, 0.91)	0.68 (0.56, 0.81)
No	-	-	XGBoost	0.84 (0.79, 0.91)	0.59 (0.48, 0.69)
Yes	AUC	0.25	Logistic regression	0.85 (0.82, 0.89)	0.56 (0.46, 0.69)
Yes	BAL	-	Logistic regression	0.70 (0.64, 0.83)	0.54 (0.43, 0.65)
Yes	F1	4.9	Logistic regression	0.85 (0.82, 0.89)	0.55 (0.46, 0.69)
Yes	AUC	0.3	Random forest	0.84 (0.77, 0.91)	0.67 (0.55, 0.80)
Yes	BAL	-	Random forest	0.84 (0.76, 0.91)	0.67 (0.56, 0.79)
Yes	F1	8.6	Random forest	0.85 (0.80, 0.91)	0.67 (0.54, 0.81)
Yes	AUC	0.05	XGBoost	0.30 (0.14, 0.50)	0.58 (0.48, 0.68)
Yes	BAL	0.321	XGBoost	0.72 (0.65, 0.79)	0.59 (0.48, 0.69)
Yes	F1	0.8	XGBoost	0.84 (0.78, 0.91)	0.59 (0.48, 0.71)

Optim: Method used to select positive class weight. AUC: Best AUC from all evaluated class weights. F1: Maximum F1 score from all evaluated class weights. BAL: Dataset distribution-based weighting. AUC, area under the receiver operating characteristic curve.

Supplementary Table 3. Test set performance for weighted training. 5-fold cross-validation bootstrap mean (95% CI) from 1000 random samplings of one series/study. Performance based on ensemble of models

Weight	Optim	Model	Accuracy	PPV	NPV	Sensitivity	Specificity	F1	AUC
Y	AUC	LR	0.81 (0.77, 0.85)	0.80 (0.77, 0.83)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	0.16 (0.00, 0.33)	0.89 (0.87, 0.91)	0.70 (0.59, 0.82)
Y	BAL	LR	0.58 (0.50, 0.65)	0.95 (0.89, 1.00)	0.34 (0.29, 0.40)	0.48 (0.40, 0.55)	0.91 (0.83, 1.00)	0.63 (0.55, 0.71)	0.72 (0.59, 0.86)
Y	F1	LR	0.81 (0.77, 0.85)	0.80 (0.77, 0.83)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	0.16 (0.00, 0.33)	0.89 (0.87, 0.91)	0.68 (0.55, 0.81)
Y	AUC	RF	0.69 (0.62, 0.77)	0.80 (0.75, 0.85)	0.33 (0.17, 0.50)	0.80 (0.70, 0.90)	0.33 (0.17, 0.50)	0.80 (0.74, 0.86)	0.61 (0.50, 0.72)
Y	BAL	RF	0.71 (0.62, 0.81)	0.81 (0.75, 0.86)	0.36 (0.17, 0.60)	0.82 (0.75, 0.90)	0.33 (0.17, 0.50)	0.81 (0.75, 0.88)	0.62 (0.52, 0.71)
Y	F1	RF	0.77 (0.73, 0.81)	0.78 (0.76, 0.80)	0.51 (0.00, 1.00)	0.97 (0.95, 1.00)	0.09 (0.00, 0.17)	0.87 (0.84, 0.89)	0.53 (0.43, 0.63)
Y	AUC	XGB	0.33 (0.31, 0.35)	1.00 (1.00, 1.00)	0.26 (0.25, 0.26)	0.13 (0.10, 0.15)	1.00 (1.00, 1.00)	0.22 (0.18, 0.26)	0.63 (0.53, 0.75)
Y	BAL	XGB	0.48 (0.42, 0.54)	0.79 (0.69, 0.90)	0.24 (0.15, 0.31)	0.45 (0.45, 0.45)	0.58 (0.33, 0.83)	0.57 (0.55, 0.60)	0.64 (0.49, 0.78)
Y	F1	XGB	0.73 (0.65, 0.81)	0.84 (0.79, 0.89)	0.43 (0.29, 0.60)	0.80 (0.70, 0.90)	0.50 (0.33, 0.67)	0.82 (0.76, 0.88)	0.64 (0.50, 0.78)

Optim: Method used to select positive class weight. AUC: Best AUC from all evaluated class weights. F1: Maximum F1 score from all evaluated class weights. BAL: Dataset distribution-based weighting. RF, random forest; LR, logistic regression; XGB, XG boost; AUC, area under the receiver operating characteristic curve; CI, confidence interval.