Prediction of Liver Regeneration Post-Radiotherapy Using Machine Learning and Deep Learning Models

Yining Zha[1], Aashish C. Gupta[1], Guillaume Cazoulat[1], Brian Sandeep De[2], Iwan Paolucci[3], Molly McCulloch[1], Mais M. Al Taie[1], Sireesha Yedururi[1], Eugene Jon Koay[2], Kristy K. Brock[1]

(1) Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA,
(2) Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA,
(3) Interventional Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center — Making Cancer History®

Background and Aims

One of the most common cause of mortality for liver cancer patients is hepatic failure caused by limited functional liver volume post therapy [1]. Accurate prediction of liver regeneration is crucial to maximize patient survival rate. However, robust models to predict functional liver volume on segment basis post-radiotherapy (RT) haven't been developed.

Our aims for the project are –

1. To identify the clinical and dosimetric factors that are responsible for regeneration of liver segments 1, 2, 3, 4, 5-8, 2-3 post-RT
2. To develop machine learning (ML) and deep learning (DL) models to predict liver regeneration post-RT
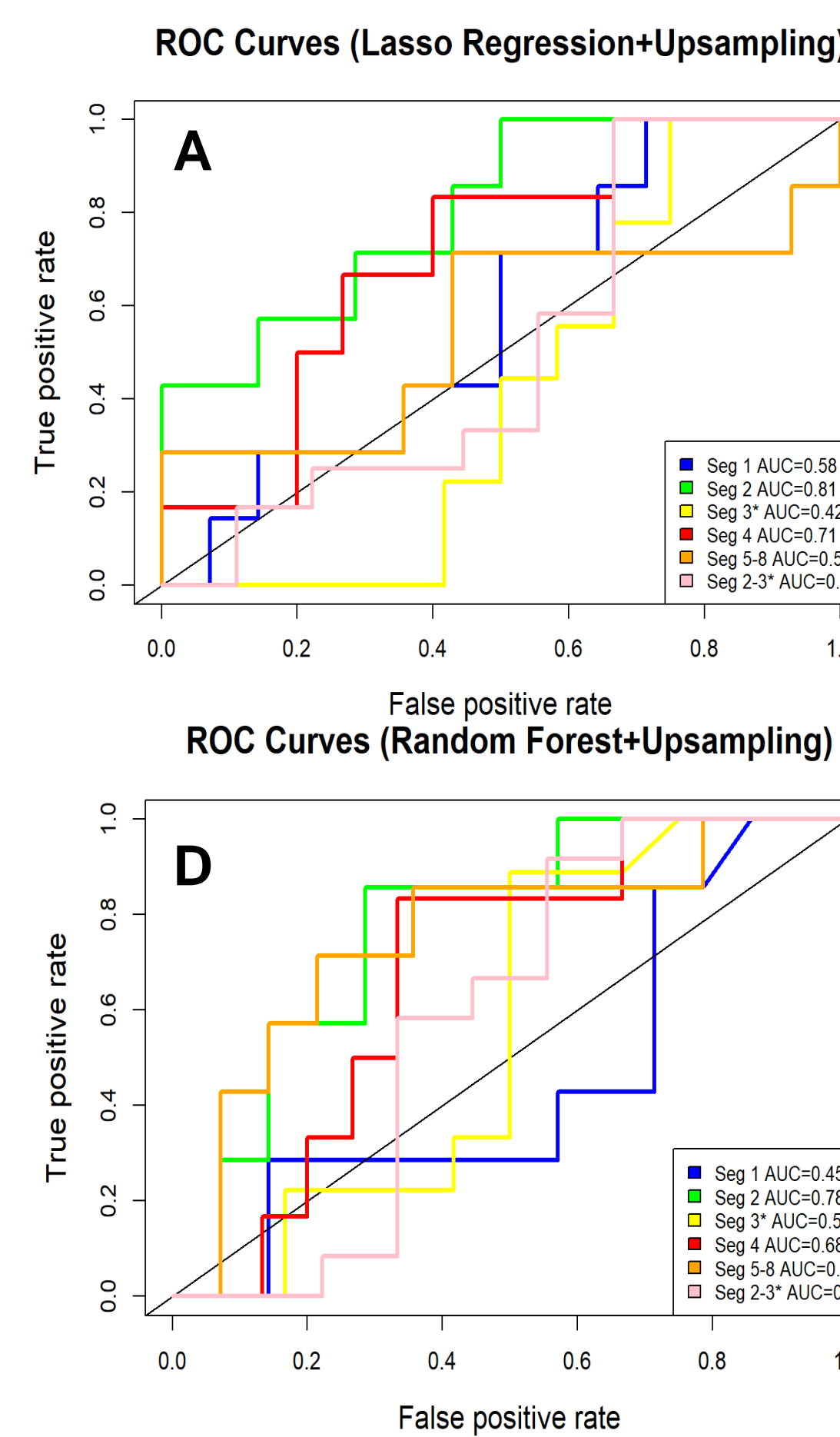
Method: Data Profile

1. Total 189 patients, 105 patients with photon radiation therapy have no missing values.
2. 30 Categorical variables, 92 Numerical variables, and 6 Response variables.
3. Binary outcome: Hypertrophy (H), Not Hypertrophy (N) for each segment.
4. Dataset size: ML: Train: 81, Test: 24
   DL: Train: 80, Test: 25

Table 1: Segmental responses at the 3-months-followup

	Seg1	Seg2	Seg3	Seg4	Seg5-8	Seg2-3
H	26	24	47	30	33	50
N	79	81	58	75	72	55
IR	3.0	3.4	1.2	2.5	2.2	1.1

H=hypertrophy, N=not hypertrophy, IR=imbalance ratio

Method: Data Exploration and Modeling

Figure. 1: Workflow of the study. Univariate and multivariable analysis were performed to identify significant variables. Resampling was performed where imbalance ratio was greater than 2 for ML and for all DL models. 10-fold Cross Validation was performed for all models. For DL, number of nodes in the hidden layer and decay rates were tuned by training models with nodes=1:1:10 and decay rate = 0.1:0.1:0.5 [2]

Resampling was performed for all models using three methods: Upsampling, Downsampling, Random Over-Sampling Examples (ROSE). ROSE was only used for Deep Learning.

Results: Data Exploration

	Variables	Odds Ratio	95% CI	p-value
Seg1	$V_{<15Gy_1}$	1.04	[1.02, 1.07]	< .001
	D_{90_4}	0.97	[0.95, 0.99]	0.002
	$D_{90_{5-8}}$	1.03	[1.01, 1.05]	0.003
	In Segment 5-8 [Yes]	0.2	[0.05, 0.72]	0.019
	D_{90_1}	0.98	[0.97, 1.00]	0.025
	$D_{Tot[80-100Gy]}$	44.45	[1.14, 2368.08]	0.045
Seg2	D_{min_3}	0.97	[0.96, 0.99]	< .001
	D_{90_4}	0.98	[0.96, 0.99]	0.004
	$V_{<20Gy_{5-8}}$	0.97	[0.95, 0.99]	0.009
	D_{95_1}	0.98	[0.97, 1.00]	0.014
Seg3	$D_{90_{2-3}}$	0.99	[0.97, 1.00]	0.02
	V_{s2}	1.01	[1.00, 1.03]	0.025
	Gender [Male]	0.34	[0.12, 0.90]	0.034
	In Segment 5-8 [Yes]	2.85	[1.03, 8.43]	0.049
Seg4	D_{95_3}	1.03	[1.01, 1.06]	0.002
	D_{min_3}	0.97	[0.95, 0.99]	0.013
	$D_{50_{5-8}}$	0.98	[0.97, 1.00]	0.013
	In Segment 4 [Yes]	0.22	[0.06, 0.72]	0.016
	V_{S4}	0.99	[0.99, 1.00]	0.018
	EQD2	0.96	[0.93, 1.00]	0.046
Seg5-8	V_{S3}	1	[0.99, 1.00]	0.007
	$D_{99_{5-8}}$	1.01	[1.00, 1.02]	0.022
	Portal Vein Thrombosis [Yes]	0.22	[0.05, 0.78]	0.031
Seg2-3	D_{min_4}	0.99	[0.97, 1.00]	0.015
	V_{S2}	1.01	[1.00, 1.02]	0.034
	D_{1_3}	0.99	[0.98, 1.00]	0.045

Table 2 highlights: Odds Ratios by Multiple Logistic Regression with Backward Selection. P-value threshold<0.05 means statistically significant. (Highlighted cells indicate results excluding CI~1)

Highlights for Seg 1 response:

- Risk increases by 4% for every 1% of volume received <15 Gy in segment 1 given all other factors controlled.
- Risk decreases by 80% for patient with tumor in segment 5-8 given all other factors controlled.
- Total dose [80-100Gy] has both large OR and 95% CI due to small number of cases (4/105).
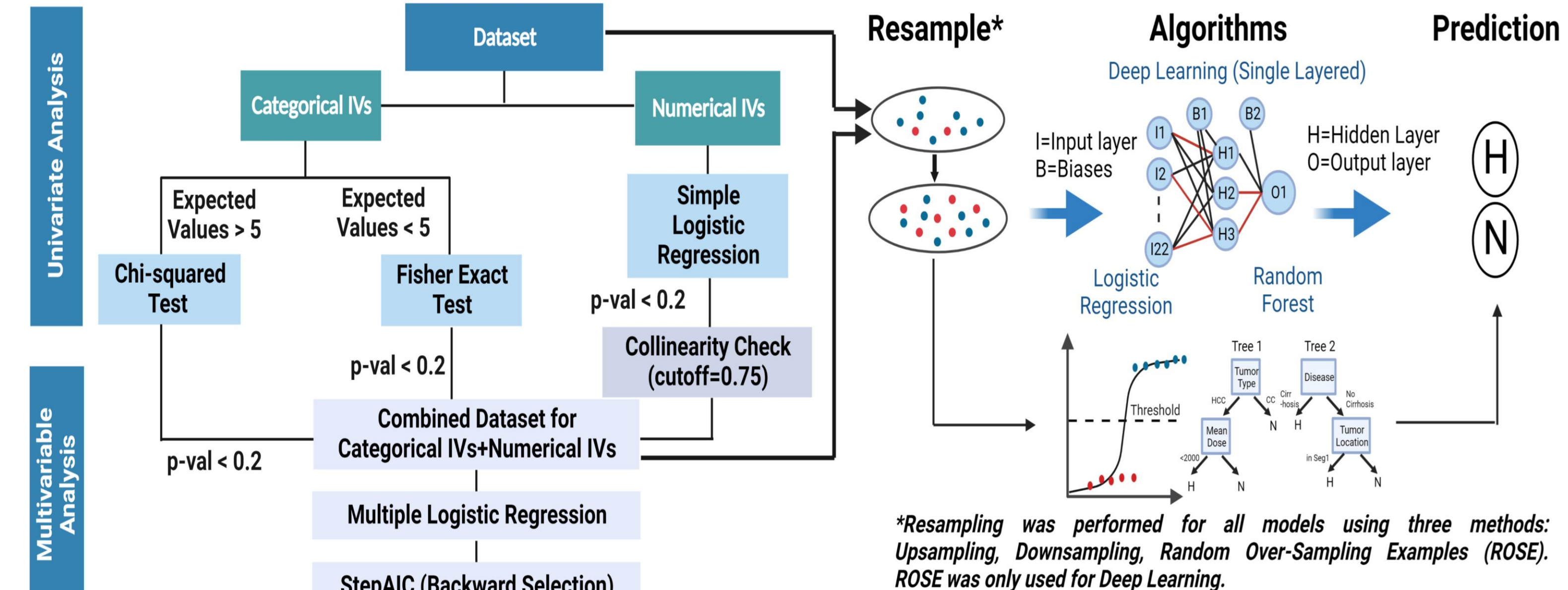
Results: Modeling

Table 4: Test Accuracy for all models.

Test Accuracy	LR+ Upsample	LR+ Downsample	LR+ ROSE	RF+ Upsample	RF+ Downsample	RF+ ROSE	DL+ ROSE
Seg1	0.57	0.57	0.57	0.62	0.38	0.48	0.70
Seg2	0.67	0.67	0.81	0.76	0.67	0.48	0.55
Seg3	0.52	0.52	0.52	0.62	0.62	0.62	0.50
Seg4	0.71	0.71	0.76	0.67	0.62	0.52	0.62
Seg5-8	0.57	0.67	0.71	0.76	0.67	0.52	0.45
Seg2-3	0.48	0.48	0.48	0.57	0.57	0.57	0.43

Figure. 2: ROC curves for various ML models and DL models with different resampling methods are shown in **A-F** and in **G**, respectively. LR + ROSE and RF + Upsampling have overall highest test accuracies for all six segments. RF models have better performance with larger dataset. All random forest models have high train accuracy 100% but relative low test accuracy, which indicates overfitting. LR is suggested to use for prediction because it has no overfitting issues. DL model showed superior results than ML for segment 1.
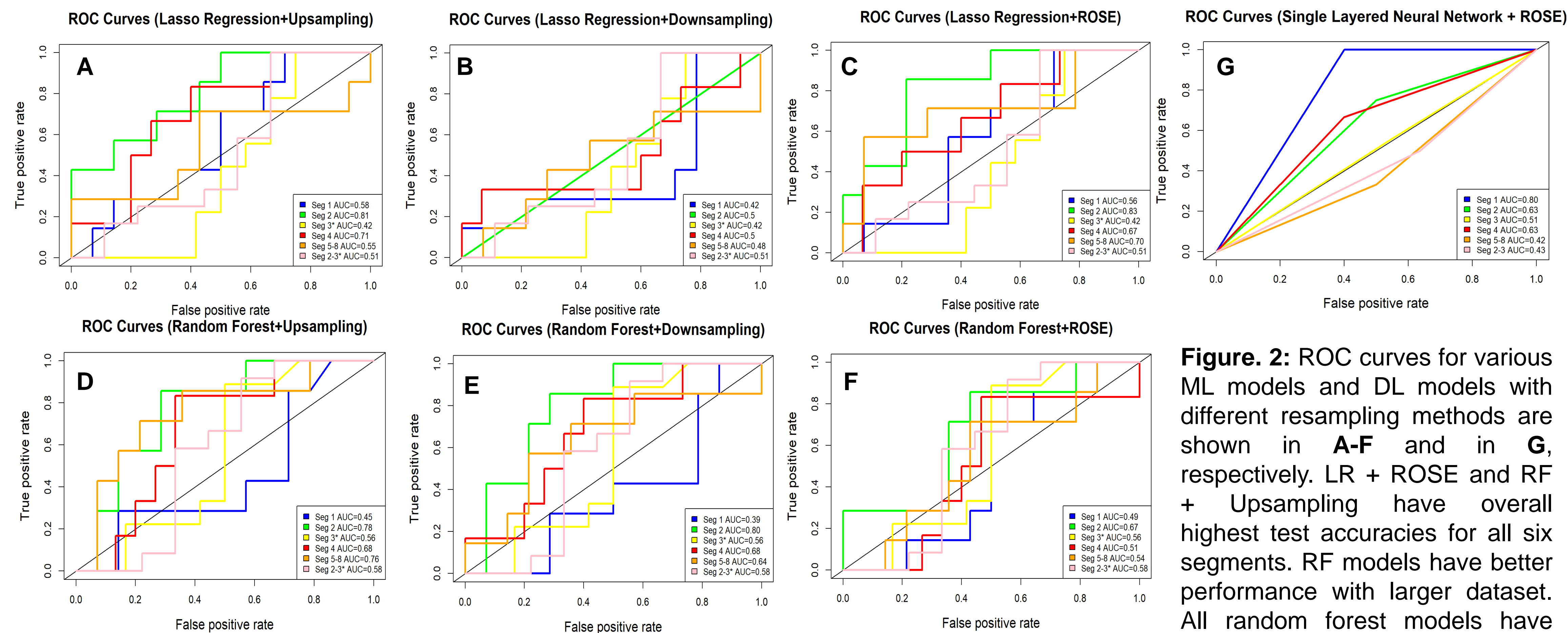
Conclusions

1. We developed ML and DL models with three resampling methods to predict hypertrophy.
2. The multivariable analysis results (Table 2) demonstrates that tumor locations and dosimetric variables are significant protective/risk factors for liver hypertrophy.
3. Tumor locations and dosimetric variables are important predictors for all ML and DL models.
4. Most models have higher sensitivity than specificity. Though test accuracies are low, our models are still useful to predict hypertrophy cases.
5. Overall, ML models showed superior results than DL models.

References

1. Kong FH, Miao XY, Zou H, Xiong L, Wen Y, Chen B, Liu X, Zhou JJ. End-stage liver disease score and future liver remnant volume predict post-hepatectomy liver failure in hepatocellular carcinoma. World J Clin Cases. 2019 Nov 26;7(22):3734-3741. doi: 10.12998/wjcc.v7.i22.3734.
2. Adapted from BioRender.com (2022). Retrieved from https://app.biorender.com/illustrations/62e6fcb1f2dd732bb682460b

yzha@mdanderson.org
kkbrock@mdanderson.org