

Online Table 1: Various test metrics for regression. Performance reflects mean of 5 separate seeded runs

Regression	Accuracy	RMSE	Sensitivity	Specificity	PPV	NPV	AUC ROC
1 section, random initialization	85.66% (2.15%)	0.1078 (0.0142)	90.20% (1.95%)	81.05% (2.81%)	82.69% (3.10%)	89.23% (1.60%)	0.9468 (0.0119)
1 section, pretrained	93.84% (0.95%)	0.0594 (0.0047)	93.88% (1.80%)	93.76% (0.75%)	93.81% (0.72%)	93.86% (1.69%)	0.9868 (0.0066)
3 sections, random initialization	86.94% (1.60%)	0.0972 (0.0109)	90.91% (1.44%)	82.87% (2.38%)	84.23% (2.23%)	90.10% (1.15%)	0.9542 (0.0070)
3 sections, pretrained	95.12% (1.11%)	0.0539 (0.0036)	95.12% (1.74%)	95.08% (1.19%)	95.14% (1.19%)	95.11% (1.67%)	0.9891 (0.0054)
9 sections, random initialization	87.67% (2.88%)	0.0883 (0.0038)	92.10% (2.15%)	83.32% (4.31%)	84.66% (4.28%)	91.29% (2.19%)	0.9581 (0.0163)
9 sections, pretrained	95.12% (1.32%)	0.0465 (0.0047)	94.63% (1.12%)	95.58% (2.18%)	95.57% (2.08%)	94.69% (0.84%)	0.9907 (0.0042)
27 sections, random initialization	87.17% (1.10%)	0.0918 (0.0047)	92.19% (1.96%)	82.08% (1.93%)	83.83% (1.77%)	91.29% (1.91%)	0.9572 (0.0095)
27 sections, pretrained	94.88% (1.45%)	0.0487 (0.0067)	94.18% (1.27%)	95.59% (1.98%)	95.52% (2.06%)	94.24% (1.13%)	0.9911 (0.0037)

The numbers in parentheses represent SD.

Online Table 2: Various test metrics for binary classification on ResNet-152. Performance reflects mean of 5 separate seeded runs

Binary Classification	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC ROC
1 section, random initialization	90.78% (1.04%)	88.47% (1.59%)	93.05% (1.23%)	92.79% (1.24%)	88.90% (1.60%)	0.9646 (0.0092)
1 section, pretrained	92.71% (0.72%)	90.54% (2.03%)	94.77% (1.94%)	94.78% (1.41%)	90.92% (1.59%)	0.9785 (0.0054)
3 sections, random initialization	92.29% (0.64%)	90.79% (1.22%)	93.74% (1.92%)	93.68% (1.60%)	91.00% (1.09%)	0.9718 (0.0053)
3 sections pretrained	93.49% (1.46%)	91.24% (2.52%)	95.81% (0.87%)	95.60% (1.03%)	91.52% (2.76%)	0.9856 (0.0042)

The numbers in parentheses represent SD.

Online Table 3: Various test metrics for regression on ResNet-152. Performance reflects mean of 5 separate seeded runs

Regression	Accuracy	Test RMSE	Sensitivity	Specificity	PPV	NPV	AUC ROC
1 slice, random initialization	84.96% (1.93%)	0.1556 (0.0703)	88.92% (1.36%)	81.04% (3.79%)	82.47% (3.80%)	87.93% (0.90%)	0.9333 (0.0106)
1 slice, pretrained	94.46% (1.16%)	0.0603 (0.0054)	94.01% (2.80%)	94.84% (1.10%)	94.88% (0.87%)	94.14% (2.43%)	0.9875 (0.0066)
3 slices, random initialization	87.95% (1.83%)	0.1205 (0.0564)	89.28% (1.79%)	86.59% (2.56%)	86.94% (2.87%)	88.98% (1.31%)	0.9540 (0.0132)
3 slices, pretrained	95.16% (0.88%)	0.0536 (0.0045)	95.04% (1.89%)	95.24% (0.33%)	95.27% (0.24%)	95.06% (1.83%)	0.9897 (0.0056)

Note:—RMSE indicates root mean-squared error. The numbers in parentheses represent SD.

Online Table 4: Details of linear mixed-effects models related to number of input slices^a

Slices	Accuracy		Sensitivity		Specificity	
	Odds Ratio	P Value	Odds Ratio	P Value	Odds Ratio	P Value
3	1.10 (0.99–1.21)	.077	1.12 (0.96–1.30)	.158	1.10 (0.94–1.28)	.222
9	1.13 (1.02–1.25)	.022	1.20 (1.02–1.40)	.024	1.09 (0.94–1.28)	.254
27	1.14 (1.03–1.27)	.013	1.19 (1.02–1.39)	.027	1.12 (0.96–1.31)	.144
Comparisons						
3 vs. 9 slices		.608		.398		.936
3 vs. 27 slices		.482		.421		.810
9 vs. 27 slices		.849		.968		.748

^a All odds ratios are compared with 1 section input.