

Multicentre performance and consistency of two deep learning models for malignancy probability estimation of incidental pulmonary nodules

ELECTRONIC SUPPLEMENTARY MATERIAL

Supplemental 1

Table S1: Maximum nodule size stratified by centre and malignancy status per size bin

	<i>Benign</i>	<i>Malignant</i>	<i>p</i>	<i>Overall</i>	<i>N (benign/ malignant)</i>
All sizes					
Centre 1	11.7 [9.5, 16.3]	16.4 [13.6, 23.4]	<0.01	13.8 [10.1, 18.8]	60/29
Centre 2	12.6 [9.4, 18.8]	13.1 [10.0, 18.4]	0.57	12.8 [9.4, 18.8]	57/30
Centre 3	12.1 [9.1, 17.5]	12.5 [9.6, 20.1]	0.99	12.2 [9.3, 19.1]	60/30
5-10mm					
Centre 1	8.4 [6.4, 9.6]	9.3 [8.1, 9.6]	0.69	8.4 [6.5, 9.7]	21/3
Centre 2	8.9 [7.6, 9.4]	9.0 [7.0, 9.3]	0.92	8.9 [7.5, 9.4]	20/9
Centre 3	8.0 [6.8, 9.0]	8.1 [6.6, 9.5]	0.76	8.0 [6.7, 9.2]	20/10
10-15mm					
Centre 1	11.9 [11.2, 14.1]	13.9 [12.5, 14.7]	0.10	12.6 [11.5, 14.2]	22/10
Centre 2	13.0 [12.0, 14.0]	12.9 [12.5, 13.9]	0.56	12.9 [12.0, 14.0]	20/11
Centre 3	12.1 [11.5, 13.9]	12.5 [11.4, 13.5]	0.86	12.2 [11.4, 13.7]	20/10
15-30mm					
Centre 1	19.1 [17.8, 21.4]	23.4 [19.6, 26.2]	0.04	20.4 [18.1, 23.9]	17/16
Centre 2	20.5 [19.7, 24.6]	22.2 [19.3, 24.3]	0.96	20.9 [19.3, 24.6]	17/10
Centre 3	21.8 [17.8, 26.8]	22.0 [20.3, 22.9]	0.89	22.0 [19.5, 26.5]	20/10

Supplemental 2

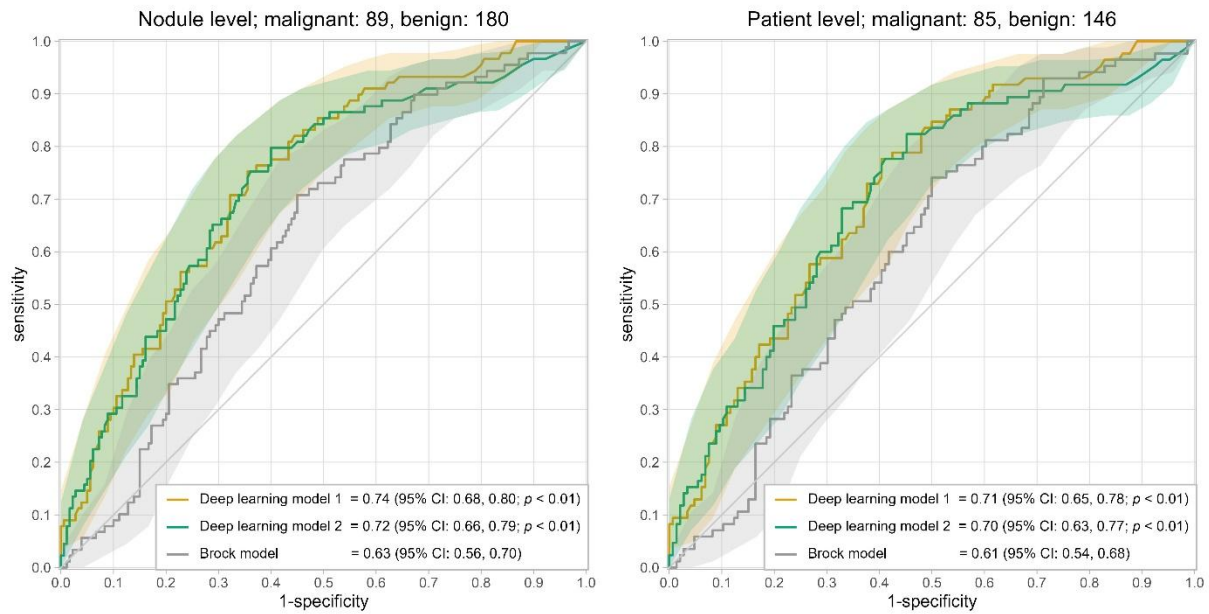


Figure S2: Sensitivity analysis comparing ROC curves at nodule level (as presented in Figure 3) and patient level, showing ROCs with 95% confidence bands and AUCs with 95% confidence intervals for the malignancy risk models, selected as the highest risk score. *P*-values correspond to comparisons with the Brock model.