

**Supplementary Table 3. Lesion-Based Diagnostic Performance of DLLD and Readers in the Binary Classification (Table 3) \***

Testee	Sensitivity (%) (95% CI)	DLLD vs. Reader ( <i>P</i> )	FPP (95% CI)	DLLD vs. Reader ( <i>P</i> )
DLLD	69.70 [69/99] (60.64, 78.75)		0.541 [46/85] (0.433, 0.649)	
Radiologist		0.750		< 0.001 <sup>†</sup>
Reader-averaged	68.01 (62.68, 73.35)		0.126 (0.085, 0.166)	
Reader 1	68.69 [68/99]		0.071 [6/85]	
Reader 2	61.62 [61/99]		0.047 [4/85]	
Reader 3	73.74 [73/99]		0.259 [22/85]	
Resident		0.332		< 0.001 <sup>†</sup>
Reader-averaged	64.31 (58.83, 69.79)		0.247 (0.194, 0.300)	
Reader 4	73.74 [73/99]		0.471 [40/85]	
Reader 5	70.71 [70/99]		0.224 [19/85]	
Reader 6	48.48 [48/99]		0.047 [4/85]	

\*Performance comparison when the internal confidence percent of the algorithm was 80% and 4 points of the readers were used as the threshold for positive judgment, <sup>†</sup>Statistically significant, CI = confidence interval, DLLD = deep learning-based lesion detection algorithm, FPP = false positives per patient