

**Supplementary Table 4. 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            | 37.37 [37/99] (27.84, 46.90) |                              | 0.094 [8/85] (0.032, 0.156) |                              |
| Radiologist     |                              | 0.068                        |                             | 0.109                        |
| Reader-averaged | 47.81 (42.10, 53.53)         |                              | 0.039 (0.015, 0.063)        |                              |
| Reader 1        | 50.51 [50/99]                |                              | 0.035 [3/85]                |                              |
| Reader 2        | 38.38 [38/99]                |                              | 0.024 [2/85]                |                              |
| Reader 3        | 54.55 [54/99]                |                              | 0.059 [5/85]                |                              |
| Resident        |                              | 0.951                        |                             | 0.384                        |
| Reader-averaged | 37.04 (31.51, 42.56)         |                              | 0.063 (0.033, 0.093)        |                              |
| Reader 4        | 63.64 [63/99]                |                              | 0.118 [10/85]               |                              |
| Reader 5        | 37.37 [37/99]                |                              | 0.059 [5/85]                |                              |
| Reader 6        | 10.10 [10/99]                |                              | 0.012 [1/85]                |                              |

\*Performance comparison when the internal confidence percent of the algorithm was 100% and 5 points of the readers were used as the threshold for positive judgment. CI = confidence interval, DLLD = deep learning-based lesion detection algorithm, FPP = false positives per patient