

SUPPLEMENTARY MATERIALS 5

Lesion-Based Performance Comparison between 3-Class and 1-Class DLLDs

Three class-lesion training (metastasis, cyst, and hemangioma) and one class-lesion training (metastasis) were performed to compare the effects of the number of lesion classes on metastasis detection performance.

In the performance comparison between the two types of DLLDs for confidence score, the 3-class DLLD achieved a higher the area under the alternative free-response receiver operating characteristic curve (AUAFROC) value than the 1-class DLLD with statistical significance (3-class DLLD, 0.631, confidence interval [CI] [0.520, 0.737]; 1-class DLLD, 0.432, CI [0.346, 0.524]; $p < 0.001$).

In the comparison between the two DLLDs based on binary classification, the three-class DLLD showed a significantly lower FPP than the one-class DLLD (three-class DLLD, 1.330, CI [1.052, 1.681]; one-class DLLD, 3.365, CI [2.748, 4.120], $p < 0.001$), and no statistically significant difference in sensitivity compared to the one-class DLLD (3-class DLLD, 81.82%, CI [72.68, 88.39]; one-class DLLD, 85.86%, CI [78.48, 91.00], $p = 0.40$; Supplementary Table 1). Therefore, in this study, all the performance comparisons between DLLD and the readers were performed using the three-class DLLD.

The number of false positives that occurred during the metastasis detection can be significantly reduced using three-class (metastasis, cyst, and hemangioma) CT images for DLLD training instead of one-class (metastasis) CT images. Technically, it is possible to use only metastatic images for training DLLD for metastasis detection. However, two types of benign lesions, which are commonly observed in the liver at high incidence, were added for training to teach the differences between different types of lesions. Consequently, the FPP of DLLD could be significantly reduced without reducing the sensitivity of DLLD. This multiclass training method is expected to aid the field of medical imaging, where obtain a sufficient number of training images is relatively difficult [11,12], since it can effectively reduce the number of false positives. For designing and using the multiclass training method, specific knowledge in radiology is needed to list the types of lesions with morphological similarity and high coincidence with the target lesion [13,14].