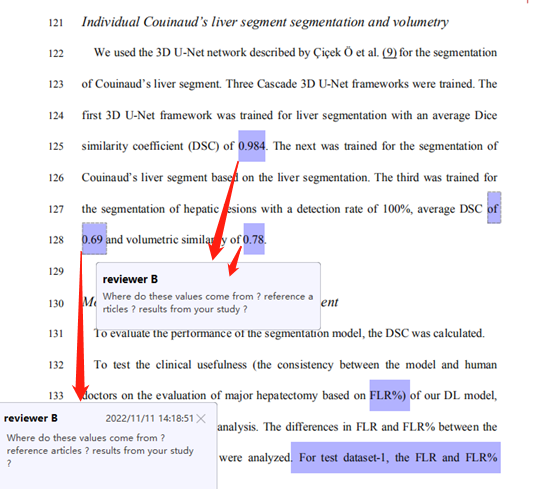
*Replies about DL model of liver segmentation and hepatic lesion*

Questions from reviewer B were presented as below:  


**Reply:**

Three cascade 3D U-Net frameworks had been trained in our institution for the automated segmentation of liver, Couinaud’s liver segment and hepatic lesion, but the results had not been reported yet.

It’s not our object to develop different DL models. What we really interested in was that how these DL models performed in real clinical practice, and how these models can affect the workflow of a surgeon in surgical planning. We prefer to evaluate these models as a whole in the workflow of clinical practice, rather than evaluate and report these DL models one by one. So, we haven’t reported the performance of these DL models individually. Here, we show the performance of the other two DL models for you:

|  |  |  |
| --- | --- | --- |
|  | **DL model for  liver segmentation** | **DL model for  hepatic Lesion segmentation** |
| ***Performance*** |  |  |
| **DSC** | **0.984 ± 0.011** | **0.693 ± 0.247** |
| **Volumetric similarity** | **0.998 ± 0.010** | **0.784 ± 0.213** |
| ***Cohort (CT scans/Patients)*** |  |  |
| **Train** | **2992/650** | **492/492** |
| **External validation** | **178/178** | **32/32** |

Segmentation results are shown below:  


Reply for subgroup analysis of DSC in test dataset 1+2  
  
The DSC differences between S1 and S3, S1 and S4, S1 and S5, S1 and S6, S1 and S7, S1 and S8, S2 and S3, S2 and S4, S2 and S5, S5 and S7, S5 and S8 were statistically significant (*P* < 0.05). No differences were found between other two groups.