Clinical validation of the deep learning-based pre-operative auxiliary planning system for liver cancer: A randomised controlled trial

Dawei Wang[1], Zhaobi Zhu[2], Fule Wu[1], Chen Xia[1], Shaokang Wang[1], Isaac Ajayi[1]

InferVision[1], The Second Affiliated Hospital of Soochow University[2]

Purpose

To evaluate the effectiveness of the deep learning (DL)-based fully automatic pre-operative planning system for liver cancer surgery.

Methods and materials

40 patients who underwent pre-operative abdominal CT and liver surgery were retrospectively enrolled. Two attending surgeons and two resident surgeons were randomly divided into control (manual 3D reconstruction using semi-automated software for surgery planning) and experimental (surgery planning based on the automatically generated reconstruction results with or without minor adjustments) groups. Primary endpoints include the anatomical structure identification accuracy and time consumption for surgery planning. Gold standard for anatomical structures determined by an experienced chief surgeon by referencing peri-operative information, including abdominal CT images, surgical records, case reports, surgical videos, etc. with ≥1cm safe margin and 80% confidence level set for statistical analyses. Statistical analysis of superiority carried out at a one-sided significance level of 0.025 for the primary endpoint of time consumption, and other indicators at a two-sided significance level of 0.05 using R statistical software.

Results

For primary endpoints, accuracy of target portal vein branch roots, reconstructed portal vein variations, reconstructed hepatic artery variations, and reconstructed hepatic artery variation was 67.50% (95%CI: 50.87%, 81.43%), 87.50% (95%CI: 73.20%, 95.81%), 77.50% (95% CI : 61.55%, 89.16%), and 87.50% (95%CI: 73.20%, 95.81%) on experimental group, which were non-inferiority at a set non-inferiority margin δ of 0.15 to the control group whose accuracy was 61.25% (95%CI: 44.56%, 76.21%), 80.00% (95%CI: 64.35%, 90.95%), 61.25% (95%CI: 44.56%, 76.21%), and 80.00% (95%CI: 64.35%, 90.95%). Average time consumption in the experimental group decreased by about 26 minutes, and the lower limit of the 95% CI was much greater than the superiority margin of 0, achieving superiority over the control group. For secondary endpoints, accuracy of target hepatic artery (68.75%, 95%CI: 52.16%, 82.44% vs 61.25%, 95%CI: 44.56%, 76.21%) and target hepatic vein (70.00%, 95%CI: 53.47%, 83.44% vs. 56.25%, 95%CI: 39.69%, 71.85%), and selection accuracy of surgical methods (78.75%, 95%CI: 62.94%, 90.06% vs. 73.75%, 95%CI: 57.45%, 86.36%) all proved non-inferior to the control group. As for the residual liver ratio, the intraclass correlation coefficient between the two groups reached 91.8% (95%CI: 85.1%, 95.5%). An increased software user satisfaction of 27% was observed in the experimental group.

Table 1. Primary endpoints results in both groups

Order	Primary Endpoints	Control group	Experimental group	Outcome
1	Target portal vein branch roots (Identification accuracy)	67.50% (95%CI: 50.87%, 81.43%)	61.25% (95%CI: 44.56%, 76.21%)	Non-inferiority
2	Reconstructed portal vein variations (Identification accuracy)	87.50% (95%CI: 73.20%, 95.81%)	80.00% (95%CI: 64.35%, 90.95%)	Non-inferiority
3	Reconstructed proper hepatic artery variations (Identification accuracy)	77.50% (95%CI: 61.55%, 89.16%),	61.25% (95%CI: 44.56%, 76.21%)	Non-inferiority
4	Reconstructed hepatic artery variation (Identification accuracy)	87.50% (95%CI: 73.20%, 95.81%)	80.00% (95%CI: 64.35%, 90.95%)	Non-inferiority
5	Surgery planning time-consumption	7'56''	34'11''	Superiority

Table 2. Secondary endpoints results in both groups

Order	Secondary Endpoints	Control group	Experimental group	Outcome
1	Target hepatic artery (Identification accuracy)	68.75% (95%CI: 52.16%, 82.44%)	61.25% (95%CI: 44.56%, 76.21%)	Non-inferiority
2	Target hepatic vein (Identification accuracy)	70.00% (95%CI: 53.47%, 83.44%)	56.25% (95%CI: 39.69%, 71.85%)	Non-inferiority
3	Surgical types (Selection accuracy)	78.75% (95%CI: 62.94%, 90.06%)	73.75% (95%CI: 57.45%, 86.36%)	Non-inferiority
4	Software user satisfaction (5-point Likert Scale)	97.5%	70.5%	Superiority

Conclusions

The DL-based system is non-inferior to the clinically utilised semi-automated software in aiding accurate surgery plans with surgery planning efficiency and user satisfaction significantly improved. The user-friendly and highly efficient DL-based pre-operative planning system makes making accurate surgery plans for each patient in clinical practice more convenient.

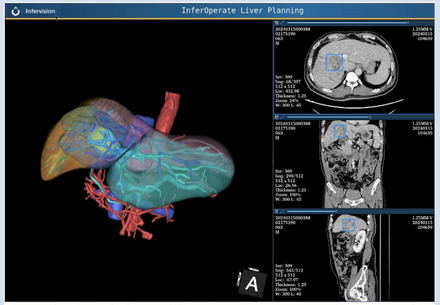


Figure 1. The interface of the deep learning-based pre-operative auxiliary planning system for liver cancer

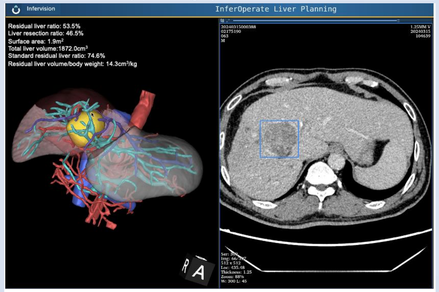


Figure 2. Simulated resection of liver lesions in the deep learning-based pre-operative auxiliary planning system