European Journal of Radiology 193 (2025) 112415

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Fully automatic bile duct segmentation in magnetic resonance cholangiopancreatography for biliary surgery planning using deep learning

Haisu Tao [a,b,c,d,e,1], Junfeng Wang [a,b,c,d,1], Kangwei Guo [a,b,c,d,1], Wang Luo [f], Xiaojun Zeng [a,b,c,d], Mingjun Lu [g], Jinyu Lin [a,b,c,d], Baihong Li [a,b,c,d], Yinling Qian [h,*], Jian Yang [a,b,c,d,*]

[a] Department of Hepatobiliary Surgery I, General Surgery Center, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China
[b] Guangdong Provincial Clinical and Engineering Center of Digital Medicine, Guangzhou 510280, China
[c] Liver Cancer Center of Zhujiang Hospital, Guangzhou 510280, China
[d] South China Institute of National Engineering Research Center of Innovation and Application of Minimally Invasive Instruments, Guangzhou 510280, China
[e] Special Medical Service Center, Zhujiang Hospital, Southern Medical University, Guangzhou 510282 China
[f] Department of Hepatobiliary-pancreatic Surgery, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou 510200, China
[g] Department of Radiology, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China
[h] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ARTICLE INFO

Keywords:
Deep learning
MRCP
Biliary tract surgery
Three-dimensional model

ABSTRACT

Objectives: To automatically and accurately perform three-dimensional reconstruction of dilated and non-dilated bile ducts based on magnetic resonance cholangiopancreatography (MRCP) data, assisting in the formulation of optimal surgical plans and guiding precise bile duct surgery.

Methods: A total of 249 consecutive patients who underwent standardized 3D-MRCP scans were randomly divided into a training cohort (n = 208) and a testing cohort (n = 41). Ground truth segmentation was manually delineated by two hepatobiliary surgeons or radiologists following industry certification procedures and reviewed by two expert-level physicians for biliary surgery planning. The deep learning semantic segmentation model was constructed using the nnU-Net framework. Model performance was assessed by comparing model predictions with ground truth segmentation as well as real surgical scenarios. The generalization of the model was tested on a dataset of 10 3D-MRCP scans from other centers, with ground truth segmentation of biliary structures.

Results: The evaluation was performed on 41 internal test sets and 10 external test sets, with mean Dice Similarity Coefficient (DSC) values of respectively 0.9403 and 0.9070. The correlation coefficient between the 3D model based on automatic segmentation predictions and the ground truth results exceeded 0.95. The 95 % limits of agreement (LoA) for biliary tract length ranged from −4.456 to 4.781, and for biliary tract volume ranged from −3.404 to 3.650 ml. Furthermore, the intraoperative Indocyanine green (ICG) fluorescence imaging and operation situation validated that this model can accurately reconstruct biliary landmarks.

Conclusion: By leveraging a deep learning algorithmic framework, an AI model can be trained to perform automatic and accurate 3D reconstructions of non-dilated bile ducts, thereby providing guidance for the preoperative planning of complex biliary surgeries.

Abbreviations: 3D, three-dimensional; MRCP, magnetic resonance cholangiopancreatography; DSC, Dice Similarity Coefficient; ICG, indocyanine green; CHD, common hepatic duct; CBD, common bile duct; LHD, left hepatic duct; RHD, right hepatic duct; RAHD, right anterior hepatic duct; RPHD, right posterior hepatic duct.

* Corresponding authors.
E-mail addresses: yl.qian@siat.ac.cn (Y. Qian), yangjian486@126.com (J. Yang).

[1] **Contributions:** Haisu Tao, Junfeng Wang and Kangwei Guo were contributed equally to this work as co-first authors.

https://doi.org/10.1016/j.ejrad.2025.112415
Received 27 November 2024; Received in revised form 1 April 2025; Accepted 10 September 2025
Available online 15 September 2025

1. Introduction

Complications from complex biliary surgery, including live donor liver transplantation, radical resection of biliary tumors, and hepatectomy for hepatolithiasis, occur at a high rate of up to about 30 % [1–4], which mainly consist of biliary injury, bleeding, bile fistula, and biliary stricture. These complications can severely impact patients' surgical outcomes, quality of life, and even endanger their lives. A key reason is the intricate and variable anatomical structure of the intrahepatic bile ducts, which are in close proximity to important intrahepatic blood vessels such as the portal vein and hepatic artery, with variable spatial relationships. Therefore, it is essential to accurately obtain individualized three-dimensional reconstruction models of the bile ducts and blood vessels before surgery, to conduct precise 3D preoperative planning, and to assist surgeons in performing precise surgical procedures.

Currently, the 3D reconstruction for hepatobiliary surgery most commonly utilizes enhanced CT imaging. Although CT data provides clear visualization of blood vessels, it poorly delineates non-dilated intrahepatic bile ducts, hindering effective biliary 3D reconstruction. To address this challenge, clinical guidelines recommend the use of magnetic resonance cholangiopancreatography (MRCP) for preoperative biliary assessment [5]. MRCP offers clear imaging of the entire biliary tree structure, regardless of duct dilation, facilitating precise three-dimensional biliary reconstruction based on MRCP data [6,7]. However, precise three-dimensional biliary reconstruction from MRCP data still largely depends on expert physicians for manual or semi-automatic segmentation [8,9], characterized by low automation, poor robustness, and high consumption of labor and time costs.

The advancement of deep learning has introduced innovative approaches for precise preoperative planning, with numerous research teams constructing 3D reconstruction models based on deep learning and achieving outstanding results [10,11]. Therefore, leveraging deep learning to build an automatic and precise 3D biliary reconstruction model is a promising method. However, few studies have been reported that utilize deep learning to build an automatic and precise 3D biliary reconstruction model based on MRCP data. Consequently, this study utilized a manually annotated MRCP dataset to train a fully automated deep learning model for the segmentation of the biliary tree and evaluated its application effectiveness in clinical practice.

2. Methods

2.1. Clinical study

This retrospective study was conducted according to the tenets of the Declaration of Helsinki and was approved by the Ethics Committee of Hospital (Ethics Approval No. 2019-JS-060-01).

2.2. Patients

This study retrospectively analyzed consecutive patients aged over 18 who underwent MRCP scanning in the Department of Hepatobiliary Surgery from September 2021 to September 2022. The exclusion criteria included: (1) severe artifacts that precluded accurate image segmentation; (2) absence of raw data necessary for retrospective image segmentation; (3) history of hepatectomy. During the study period, 314 consecutive patients met the inclusion criteria, but 65 were excluded due to severe artifacts (n = 38), lack of necessary raw data (n = 15), or a history of hepatectomy (n = 12). From the remaining patients, 208 cases were randomly divided into the training set and 41 into the internal testing set, following a 5:1 ratio (Fig. 1).

2.3. 3D-MRCP image data acquisition

The patients were scanned on an Ingenia (3.0 T, Philips Medical Systems, Netherlands) or Achieva scanner (3.0 T, Philips Medical Systems, Netherlands), both of which used a 16-channel phased-array body coil. All metal and magnetic items were removed prior to the examination. Patients were informed about the precautions and underwent breath-holding training before the examination. They also fasted for at least four hours and consumed 400 ml of Lipton Yellow Label Black tea (4 g) 15 min before the examination to minimize signal interference from fluids in the stomach and duodenum.

To localize the biliary tract, 2D axial and coronal T2-weighted half-Fourier single-shot (TSE) sequences were recorded. During the patient's breath-holding, 3D GRASE-MRCP [12] images were captured on the coronal plane with the following parameters: TR (repetition time)/TE (echo time) of 205/64 ms, a flip angle of 90°, a field of view (FOV) of 260 × 260 mm, a section thickness of 1.5–2.2 mm, 70–110 slices, an acquisition matrix of 256 × 256, a voxel resolution of 1.1 × 1.1 × 1.1 mm, and an acquisition time of 10–18 s.

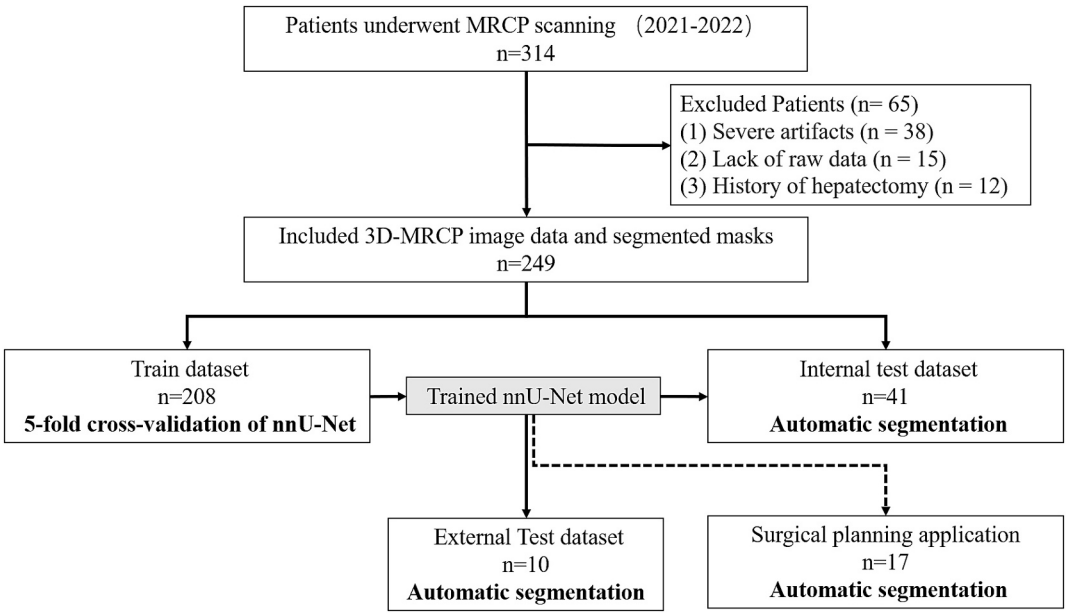

Fig. 1. Data flow of the patient selection, training, and evaluation process. MRCP: magnetic resonance cholangiopancreatography.

2.4. Ground truth segmentation

Ground truth segmentation (manual segmentation) was performed by two hepatobiliary surgeons who were familiar with biliary anatomy and trained in segmentation techniques. The original 3D-MRCP thin-layer images were manually segmented in axial and coronal views using 3D Slicer to accurately delineate the bile ducts below grade 3 branches at least. The segmented images were then reviewed by two senior radiologists and hepatobiliary surgeons. They compared the segmented images with the original images to address any issues of insufficient or excessive segmentation, ultimately establishing these as the ground truth for biliary image segmentation.

2.5. External test dataset

To the best of our knowledge, there is no publicly available dataset for 3D-MRCP biliary segmentation. Therefore, to evaluate the generalizability of our trained model, we tested it on an external dataset comprising 10 cases from Guangdong Second Provincial General Hospital. These images were acquired using an Ingenia (3.0 T, Philips Medical Systems, Netherlands) scanner with the same acquisition process. Segmentation was carried out by two hepatobiliary surgery experts from Guangdong Second Provincial General Hospital using the same methodology as described in the "Ground truth segmentation".

2.6. nnU-Net framework

nnU-Net is a deep learning network designed to adapt to any new medical image segmentation task. It automatically adjusts all parameters based on the properties of a given dataset, operating entirely without human intervention. To configure the environment for training and inference, we followed the tutorial by Isensee et al. [13] (https://github.com/MIC-DKFZ/nnUNet) (Supplemental Fig. 1). Upon receiving the images from the training dataset, the network model adaptively selected the 3D U-Net architecture, as shown in Supplemental Fig. 2, and was trained using the five-fold cross-validation method. Following the training, the model was used to make predictions on the test and external test set data. More details on the implementation of the deep learning framework are provided in Supplementary Information.

2.7. Evaluation metrics

2.7.1. Two-dimensional space evaluation

To evaluate the performance of nnU-Net in the segmentation of the test dataset and external test dataset, we used Dice Similarity Coefficient (DSC), Accuracy, Sensitivity and Specificity with expert manual annotations as the benchmark to assess the accuracy of the model on the test set. The calculation formulas are shown in Supplementary Information.

2.7.2. Three-dimensional space evaluation

In 3D Slicer, masks generated by both automatic and manual segmentation are used to reconstruct the 3D model of the bile duct. The centerline of this 3D bile duct model is then automatically extracted using a plugin to quantitatively analyze the accuracy of the segmentation results in three-dimensional space (Supplemental Fig. 3). Subsequently, the lengths of the left and right hepatic ducts, the common hepatic duct (CHD), and the common bile duct (CBD) are measured at several key confluences: the right anterior and posterior hepatic ducts (RAHD and RPHD), the bile ducts of Segment 4 (B4) and Segments 2 and 3 (B2 + 3), the left hepatic duct (LHD) and right hepatic duct (RHD), and at the duodenal papilla. In cases of anatomical variation, the length of the main trunk of the first branch of the biliary tract is measured to replace the LHD and RHD lengths. Additionally, the total volume of the entire biliary tree structure is calculated.

2.8. Surgical planning clinical application evaluation

To further validate the consistency between the 3D biliary models and patients' actual biliary structures, we preliminarily applied this approach to surgical planning for 17 cases. We retrospectively analyzed 15 patients (Non-dilated bile ducts) who underwent laparoscopic hepatectomy and 2 patients (Dilated bile ducts) underwent laparoscopic radical resection of hilar cholangiocarcinoma. The actual non-dilated biliary structure was identified intraoperatively using ICG fluorescence in hepatectomy. The consistency between the fluorescent structures and the 3D models produced by the automatic method was independently evaluated using surgical videos by a qualified hepatobiliary surgical expert. The DPM-III-01 or DPM-LIGHT-03 imaging systems (DPM, China) were used in all operations to capture ICG fluorescence images.

2.9. Statistical analysis

DSC, accuracy, sensitivity and specificity evaluation were performed in Python. Accuracy refers to the proportion of correctly classified samples to the total number of samples. Sensitivity (true positive rate) measures the ability of a model to correctly identify target regions, whereas specificity (true negative rate) assesses its ability to exclude non-target regions. (Calculation formulas are provided in the **Supplementary Information**).

Continuous data (age, weight, height) are the mean ± standard deviation (SD), while categorical variables (sex, dilated bile ducts and non-dilated bile ducts) were presented as n (%). Bile duct length and volume were presented as median and interquartile range. The Pearson correlation coefficient was chosen for evaluating correlation data acquired from automatic and manual methods. The level of agreement between automatic and manual segmentation result was assessed as described by Bland and Altman analysis (bias and 95 % limits of agreement, LoA).

Table 1
Population characteristics of the training and test sets.

	Training Set (N = 208)	Internal test Set (N = 41)	External test set (N = 10)
Age (mean ± standard deviation), years	63.7 ± 7.2	59.2 ± 8.8	55.1 ± 18.2
Sex, n (%)			
Female	97(46.66)	17(41.46)	7(70)
Male	111 (53.34)	24(58.54)	3(30)
Weight, (mean ± standard deviation), kg	63.6 ± 8.2	59.4 ± 6.8	52.3 ± 10.6
Height, (mean ± standard deviation), cm	164.3 ± 4.5	162.5 ± 6.2	160.3 ± 5.8
Dilated bile ducts, n (%)	84(40.38)	16(39.02)	8(80)
Location of obstruction, n (%)			
Intrahepatic	35(41.67)	7(43.75)	1(10)
Hilar	6(7.14)	1(6.25)	1(10)
Lower	43(51.19)	8(50)	6(60)
Type of right bile duct, n (%)			
I	126 (60.58)	25(60.98)	9(90)
II	38(18.27)	8(19.51)	0
III	15(7.21)	5(12.20)	1(10)
IV	19(9.13)	3(7.31)	0
Others	10(4.81)	0	0
Type of left bile duct, n (%)			
I	150 (72.12)	29(70.74)	8(80)
II	31(14.90)	6(14.63)	1(10)
III	27(12.98)	6(14.63)	1(10)

3. Results

3.1. Patient characteristics

Table 1 summarizes the patient characteristics of our dataset. The original 3D-MRCP images of 249 patients were included in the segmentation study of complete biliary anatomy, including 100 patients with dilated biliary tract and 149 patients with non-dilated biliary tract. The model was trained using a database from 208 patients (111 men, 97 women; mean age, 63.7 ± 7.2 years), of whom 124 patients (59.62 %) with non-dilated biliary tract underwent MRCP scanning due to health checkup, gallstones or polyps, and 84 patients (40.38 %) with dilated biliary tract had intrahepatic hepatolithiasis, hilar cholangiocarcinoma, or choledocholithiasis. The model was subsequently tested from 41 patients (24 men, 17 women; mean age, 59.2 ± 11.8 years), 25 patients with non-dilated biliary tract and 16 patients with dilated biliary tract. External testing included 10 patients (3 men, 7 women; mean age, 55.1 ± 18.2 years), including 2 patients with non-dilated biliary tract and 8 patients with dilated biliary tract (Fig. 1).

3.2. Training and automatic segmentation times

Training one epoch on the GeForce RTX 4090 took approximately 58 s (1000 epochs). For the internal test set, the average time for the manual method (ground truth) was approximately 48.96 ± 15.44 min, compared to approximately 1.80 ± 1.24 min for the automatic segmentation method. On the external test set, the manual method took an average of 55.90 ± 16.41 min, while the automatic method required

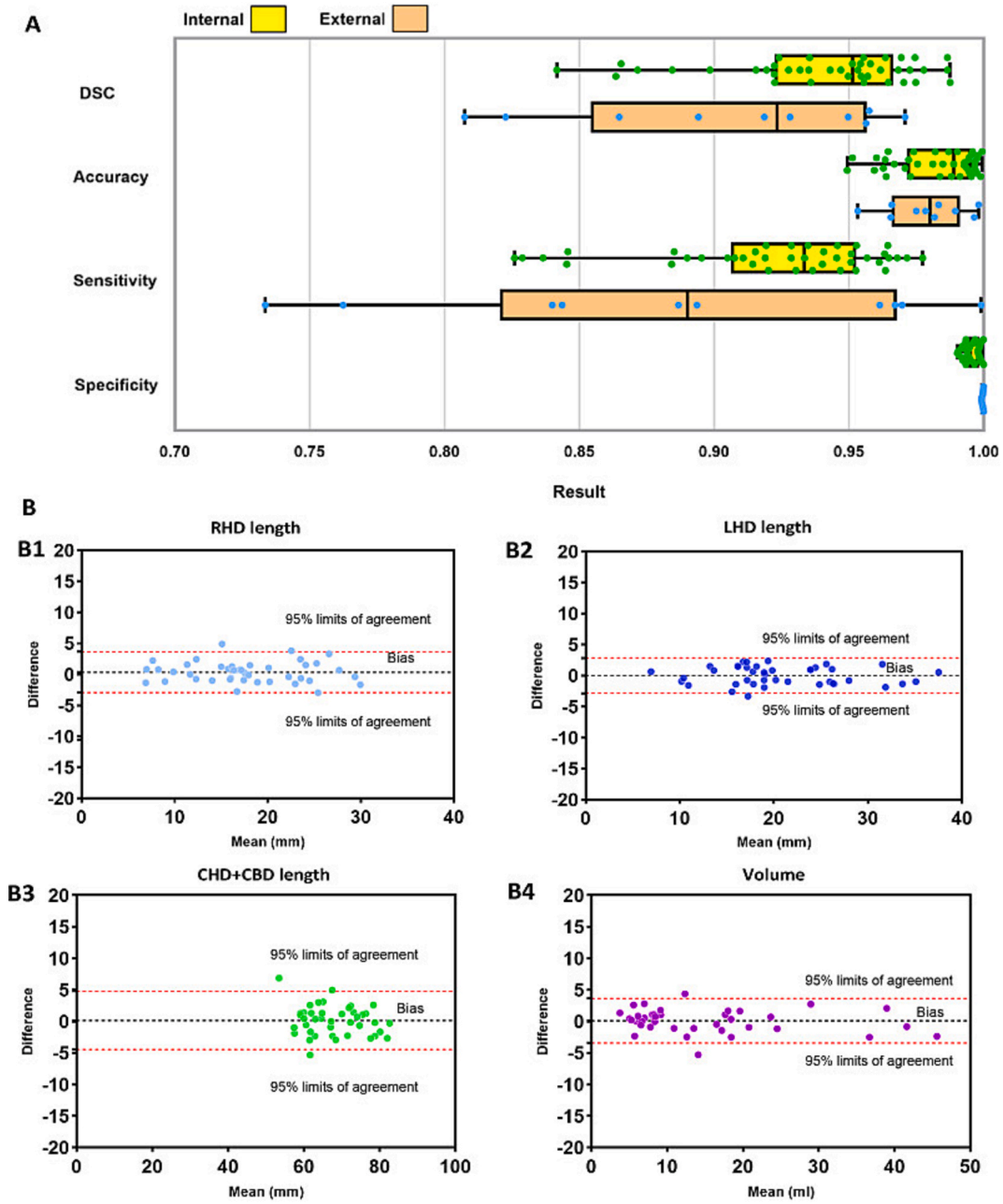

Fig. 2. (A) Distribution of DSC, Accuracy, Sensitivity and Specificity results in our test dataset. Yellow represents the internal test set and orange represents the external test set. (B) Bland-Altman plots showing agreement analysis between automatic and manual approach. (B1) RHD length, (B2) LHD length, (B3) CHD + CBD length, and (B4) total biliary tree volume. DSC: Dice Similarity Coefficient; RHD: right hepatic duct; LHD: left hepatic duct; CHD: common hepatic duct; CBD: common bile duct. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

only about 1.24 ± 0.23 min.

3.3. Evaluation results

3.3.1. Two-dimensional space segmentation evaluation

The mean Dice Similarity Coefficient (DSC) of the model on the five-fold cross-validation of 208 training sets is 0.9563. The distributions of DSC, accuracy, sensitivity, and specificity for each segmentation label are presented in Fig. 2**A**.

On the internal test set, the model achieved a mean DSC of 0.9403 (95 %CI, 0.9291 to 0.9515), with a mean accuracy of 0.9831 (95 %CI, 0.9784 to 0.9878), sensitivity of 0.9233 (95 %CI, 0.9105 to 0.9362), and specificity of 0.9955 (95 %CI, 0.9946 to 0.9964). The mean DSC for dilated ducts were 0.9527 (95 % CI: 0.9337 to 0.9717), and non-dilated ducts were 0.9327 (95 % CI: 0.9180 to 0.9474). The main segmentation effects of 5 representative cases from the internal test set are shown in Fig. 3 (representative cases 1–5).

On the external test set, the DSC, accuracy, sensitivity, and specificity results of 10 cases set are summarized in Table 2, reaching a mean DSC of 0.9070 (95 %CI, 0.8654 to 0.9486) across the entire test set, with a mean accuracy of 0.9787 (95 %CI, 0.9675 to 0.9897), sensitivity of 0.8857 (95 %CI, 0.8219 to 0.9495), and specificity of 0.9996 (95 %CI, 0.9993 to 0.9997). In the external test set, 5 cases were diagnosed with choledocholithiasis, 2 cases underwent MRCP scans for follow-up after cholecystectomy, 1 case was diagnosed with biliary stricture, and 2 cases were diagnosed with gallstones. The mean DSC for dilated ducts were 0.9228, and non-dilated ducts were 0.8437. Compared with the ground truth, only the choledochal duct below the obstruction was not segmented in 2 patients with choledocholithiasis; other major structures of the biliary tree were successfully segmented. Comparison of the predicted results with the ground truth results is shown in Fig. 3 (representative cases 6–10).

3.3.2. Three-dimensional space segmentation evaluation

The quantitative results are summarized in Table 3. For correlation analysis, the correlation coefficients (r) of 3D bile duct models constructed by the automatic and manual approaches on the internal test set were 0.9647 for right hepatic duct (RHD) length, 0.9787 for left hepatic duct (LHD) length, 0.9556 for common hepatic duct plus Common bile duct (CHD + CBD) length, and 0.9865 for total biliary tree volume (Supplemental Fig. 4). On the external test, the correlation coefficients *r* of 3D bile duct models constructed by the automatic and manual approaches were 0.9902 for RHD length, 0.9731 for LHD length, 0.9453 for CHD + CBD length, and 0.9939 for total biliary tree volume.

According to the Bland-Altman analysis, agreement between automatic and manual segmentation on the internal test set was high for RHD length (bias 0.3607 mm, 95 % LoA −2.926 to 3.648 mm), LHD length (bias 0.004 mm, 95 % LoA −2.845 to 2.853 mm), CHD + CBD length (bias 0.1622 mm, 95 % LoA −4.456 to 4.781 mm), and biliary tree volume (bias 0.1244 ml, 95 % LoA −3.404 to 3.650 ml). Bland-Altman plots are shown in Fig. 2B1-B4. On the external test, the agreement between automatic and manual segmentation was high for RHD length (bias −0.0550 mm, 95 % LoA from −1.888 to 1.778 mm), LHD length (bias −0.1250 mm, 95 % LoA from −2.487 to 2.237 mm), CHD + CBD length (bias −1.638 mm, 95 % LoA from −9.009 to 5.733 mm), and biliary tree volume (bias −0.5610 ml, 95 % LoA from −3.004 to 1.882 ml).

3.3.3. Surgical planning quantitative measurements

Based on the high accuracy and low time cost of the automatic segmentation results, we performed quantitative measurements related to surgical planning on the 3D biliary models of 259 patients.

For biliary tract classification, the type of the biliary tract in the right liver was classified according to the system proposed by Nakamura et al [14], and that of the left liver was classified based on the system proposed by Li et al [15] (Fig. 4). For the right liver, Type I was the most common, accounting for 61.78 % (160/259) of cases, followed by Types II, III, and IV, which constituted 17.76 %, 9.65 %, and 6.95 % respectively. For the left liver, Type I was predominant at 72.20 %, with Types II and III making up 14.67 % and 13.13 % respectively (Table 1).

Regarding biliary tract lengths, the average length of the right hepatic duct (RHD) was 15.56 ± 6.105 mm in 160 patients with right liver Type I. The average length of the left hepatic duct (LHD) was 18.76 ± 6.441 mm in 258 patients with left liver Types I, II, and III. The combined length of the common hepatic duct and common bile duct (CHD + CBD) averaged 71.10 ± 8.050 mm across 259 patients.

3.3.4. Surgical planning clinical application

Among the 15 patients with non-dilated biliary tracts, all biliary landmarks that were clearly visible in the MRCP data were accurately reconstructed by automatic method. All the 15 patients underwent fluorescent laparoscopic hepatectomy because of liver masses. During the operation, peripheral injection of Indocyanine green (ICG) was used to visualize the exposed biliary. Among them, 6 cases showed the

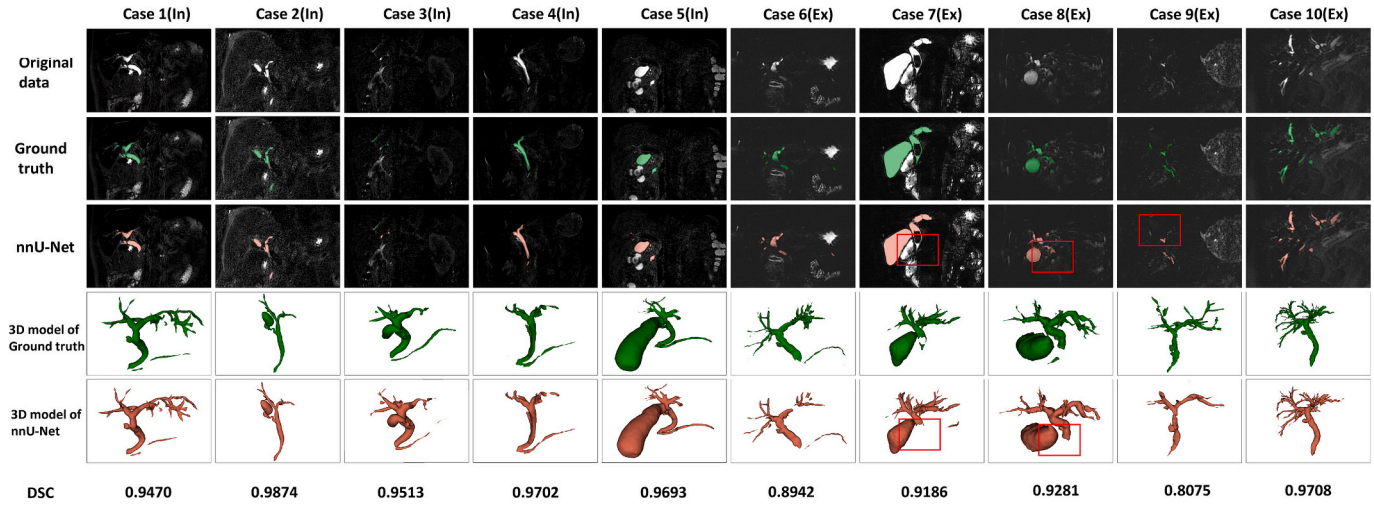

Fig. 3. 10 representative case show the raw data of the sagittal section, ground truth segmentation, nnU−Net network segmentation, 3D models reconstructed by ground truth, 3D models reconstructed by nnU-Net and the corresponding DSC results. The first 5 cases were in the internal test set, and the last 5 cases were in the external test set. Green represents the ground truth, Red represents the nnU-Net and the red box represents the part that was not segmented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Quantitative evaluation results on external test set.

Results	MRCP number									
	1	2	3	4	5	6	7	8	9	10
DSC	0.8942	0.9186	0.9281	0.8075	0.9708	0.9563	0.9497	0.8648	0.8227	0.9575
Accuracy	0.9659	0.98322	0.9749	0.9533	0.9981	0.9817	0.9894	0.9784	0.9655	0.9964
Sensitivity	0.8399	0.8436	0.8936	0.8867	0.9697	0.9990	0.9614	0.7624	0.7335	0.9671
Specificity	0.9993	0.9999	0.9998	0.9998	0.9996	0.9991	0.9996	0.9999	0.9997	0.9992

MRCP: magnetic resonance cholangiopancreatography; DSC: dice similarity coefficient.

Table 3
Quantitative analysis results.

	Automatic segmentation	Mannul segmentation
LHD, median (IQR), cm	18.98 (16.5–25.57) Range: 6.62–37.24	19.22 (16.92–25.45) Range: 7.28–37.8
RHD, median (IQR), cm	17.35 (12.66–22.44) Range: 6.54–30.78	17.51 (13.56–23.25) Range: 6.22–29.16
CHD + CBD, median (IQR), cm	67.17 (62.37–73.74) Range: 50.03–83.36	66.65 (61.74–74.35) Range: 56.58–82.51
Volume, median (IQR), cm^2	8.38 (6.67–18.25) Range: 3.08–46.77	10.35(8.39–18.33) Range: 4.54–42.02

LHD, left hepatic duct; RHD, right hepatic duct; CHD, common hepatic duct; CBD, common bile duct; IQR, interquartile range.

fluorescence of the CHD, RHD, LHD, right anterior hepatic duct (RAHD) and right posterior hepatic duct (RPHD) during right anterior hepatectomy or middle hepatic lobectomy, 1 case showed the fluorescence of the CHD, RAHD and RPHD during right posterior hepatectomy, and 7 cases showed the fluorescence of the CHD, LHD and B2/3/4 during left lateral lobectomy or hepatic segmentectomy. One case of left hemihepatectomy was excluded from the comparison results due to poor fluorescence visualization of the biliary tract. The detection rates of CHD, LHD, RHD, RAHD, RPHD, B2, B3 and B4 were 93.3 %, 86.7 %, 40.0 %, 46.67 %, 33.3 %, 40.0 %, 26.7 %, respectively. All fluorescence displays of the biliary tract structures were consistent with the 3D model (Fig. 5A–D).

For cases with dilated biliary tracts, two patients diagnosed as intrahepatic cholangiocarcinoma with hilar bile duct invasion and hilar hepatocellular carcinoma. Before operation, 3D models were reconstructed to plan the location of bile duct transection as well as to determine the distribution and number of bile duct ends post-transection. Then, both of them underwent laparoscopic radical resection of hilar cholangiocarcinoma (laparoscopic right hemi-hepatectomy plus total caudate lobectomy and biliary-enteric anastomosis). The operative video confirmed that the surgical plans for biliary tract transection were consistent with the actual outcomes (Fig. 5E and F). There were no postoperative complications such as bile leakage. Details of automatic biliary tract segmentation, surgical planning, and intraoperative situations are provided in the **Supplemental Video 1**.

Intraoperative validation using ICG biliary fluorescence and hilar cholangiocarcinoma transection planes confirmed the agreement between the predicted 3D models and actual biliary anatomy.

4. Discussion

To address the clinical need for a fully automated and precise bile duct segmentation approach, we developed and evaluated a deep learning semantic segmentation model based on a 3D neural network utilizing the nnU-Net framework. This model was specifically designed for the fully automatic segmentation of bile ducts in 3D-MRCP thin-section images. Both the training and test sets included images of dilated and non-dilated biliary tracts. The model achieved a mean Dice Similarity Coefficient (DSC) of 0.94 on the internal test set and demonstrated a high level of consistency between the automatic segmentation and ground truth, as well as between the automatic segmentation and real biliary. Finally, the high performance of nnU-net framework on an external test set proves its generalization.

In the past, surgeons could only use their experience to perform abstract 3D reconstruction and understanding of 2D images in their mind [16]. However, due to the limitations of personal clinical

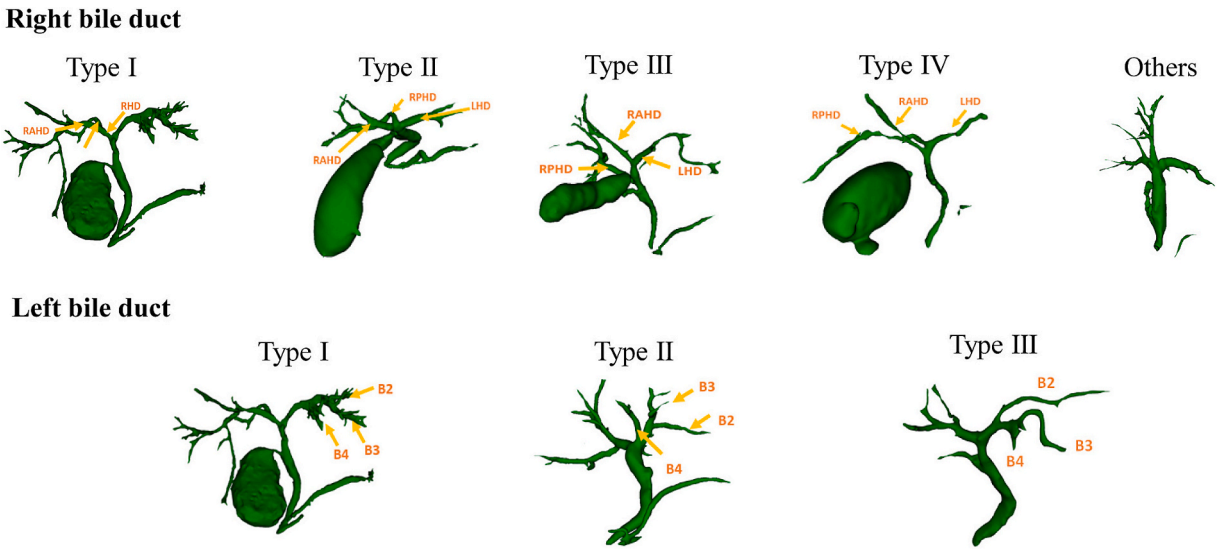

Fig. 4. Classification of bile duct based on 3D models. Right bile duct types: Type I, normal type; Type II, trifurcation of right anterior, right posterior, and left hepatic duct; Type III, early branching of right posterior duct from the common hepatic duct; Type IV, right posterior duct draining into left hepatic duct; Type V, other types of variation in bile duct anatomy. Left bile duct types: Type I, peripheral B4 joining the common trunk of B2 and B3; Type II, B4 joining the common trunk of B2 and B3 close to the hepatic hilum. Type III, B2 joining the common trunk of B3 and B4. B2, segment II duct; B3, segment III duct; B4, segment IV duct.

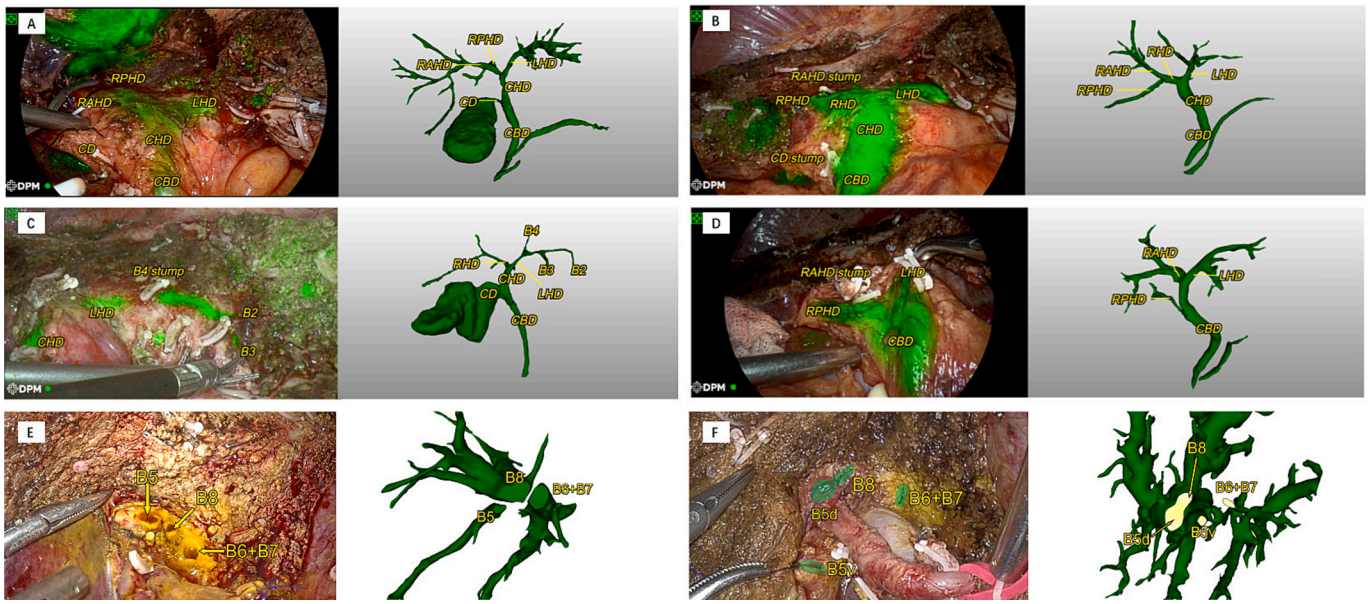

Fig. 5. Intraoperative observation of non-dilated and dilated cases via ICG fluorescence imaging (A-D) and operation situation(E-F): (A-B) Non-dilated case 1 and 2, exposing the bile ducts in the process of mesohepatectomy, comparing with 3D model. (C) Non-dilated case 3, exposing the bile ducts in the process of extended hemi-hepatectomy, comparing with 3D model. (D) Non-dilated case 4, exposing the bile ducts in the process of right anterior sectionectomy, comparing with 3D model. (E-F) Dilated case 1 and 2, exposing the end of the biliary tract after transection, comparing with surgical planning on 3D model. CD: cystic duct; CHD: common hepatic duct; CBD: common bile duct; LHD: left hepatic duct; RHD: right hepatic duct; RAHD: right anterior hepatic ducts; RPHD: right posterior hepatic ducts. B2, segment II duct; B3, segment III duct; B4, segment IV duct; B5, segment V duct; B6, segment VI duct; B7, segment VII duct; B8, segment VIII duct.

experience and abstract thinking ability, the preoperative assessment and surgical planning were often uncertain and inaccurate, especially for complex and varied intrahepatic biliary system, which are notoriously difficult to accurately evaluate, resulting in a relatively high incidence of postoperative complications [17,18]. Thus, there is a great unmet need for precise 3D models of the liver and bile ducts to assist surgical planning and quantitative analysis. 3D models of bile ducts are derived from the segmentation of medical images, include enhanced CT and MRCP. Enhanced CT utilizes contrast agents to improve the segmentation and reconstruction of blood vessels and lesions [19]. However, it performs poorly in displaying non-dilated bile ducts. Even when partial information on dilated bile ducts can be obtained, it is challenging to segment the overall biliary structures. 3D-MRCP thin-slice images offer advantages such as fast scanning speed, good continuity of tube structure, high resolution [20,21]. But commonly used methods for segmentation of bile ducts based on 3D-MRCP thin-section images are all manual or semi-automatic [8,9,22]. Although manual segmentation has always been considered the “gold standard”, it is inherently time-consuming. Namkee Oh et al. have reported using a 3D residual U-Net model for automatic segmentation of living liver donor 3D-MRCP images firstly, achieving an overall DSC of 0.80 [23]. Compared to their approach, our nnU-Net model not only achieves a higher DSC but also features a self-configuring architecture that requires no manual hyperparameter tuning, enabling surgeons to deploy the model directly. Moreover, we have extended its application to pathologically dilated bile ducts. The 3D-MRCP acquisition process and automatic segmentation algorithm model built in this study will greatly aid young surgeons to quickly and accurately understand and quantitatively analyze bile duct anatomy, contributing to the homogenization and standardization of surgical planning.

Our results demonstrated that the automatic model based on the nnU-Net framework achieved similar accuracy to manual segmentation. In spite of the small size of our MRCP training dataset and high accuracy requirements for segmentation in clinical practice, the unique structure of nnU-Net enables it to extract and learn features from MRCP images to a greater extent, achieving better segmentation accuracy using a small training set. The 3D U-Net included in nnU-Net is based on the U-Net framework, which extract contextual information from adjacent images and gradually improve the model’s ability to perceive details such as bile duct boundaries, contrast, and textural information. This may be a reason why this automatic segmentation algorithm can accurately segment the bile ducts. Another advantage of this automatic over manual is that it can achieve similar accuracy in a shorter time, which is also a key point to consider in clinical practice, especially since higher accuracy usually requires more time. When labelled by experienced experts with sufficient caution, manual segmentation can be treated as ground truth. However, such manual segmentation often takes about an hour in our result, and in clinical practice, it is difficult to ensure that every patient can obtain sufficient human resources and time allocation to complete precise bile duct segmentation. Therefore, this automatic model may help meet the clinical needs of more patients.

In addition, we attempted to use the 3D models by automated methods for clinical application. ICG is a near-infrared fluorescent dye that is almost entirely absorbed by the liver and excreted into the bile [24]. ICG fluorescence imaging can clearly and intuitively display the intra- and extrahepatic bile ducts during surgery, solving the problem that non-dilated biliary tracts are difficult to distinguish under ordinary white light. In this study, we used ICG fluorescence imaging data to validate that this model can accurately third-order branch bile duct. For complex biliary surgery, the 3D model can be observed qualitatively and quantitatively by surgeons on personal computer more freely and from multiple angles, which used to predict the location of the bile duct transection, as well as to predict the location and number of anastomoses. It is to expected to avoid the occurrence of postoperative bile leakage caused by missing some bile ducts.

Despite its highly encouraging results, our study has several limitations. First, in the prediction results of the external test set, the common bile duct below the obstruction site was not segmented in two cases of choledocholithiasis. We plan to address this issue in future work by enhancing the network framework or increasing the representation of such diseases in the training set. Additionally, while intraoperative ICG fluorescence imaging provides a visual display of the exposed bile ducts, its effectiveness is subjective and depends on the expertise of the surgeon. Furthermore, 3D models of bile ducts based on MRCP data can

only display biliary anatomy, but biliary surgery planning also requires mastery of vascular information, such as hepatic arteries, portal veins and hepatic veins. Precise 3D vascular models are usually reconstructed based on CT thin-section images. How to register and fuse bile ducts and blood vessels reconstructed from different modal data is a topic worthy of in-depth research.

In this study, we constructed a fully automated accurate bile duct segmentation algorithm based on the nnU-Net framework, an adaptive convolutional neural network (CNN) framework and state-of-the-art architecture for automatic medical image segmentation. In terms of image segmentation accuracy, this automatic segmentation algorithm can reach a level comparable to expert human performance, but its segmentation process only takes about 2 mins, enabling standardized and fast 3D segmentation of bile ducts, which is conducive to clinical application and promotion.

CRediT authorship contribution statement

Haisu Tao: Writing – review & editing, Methodology, Conceptualization, Supervision, Funding acquisition. **Junfeng Wang:** Writing – review & editing, Visualization, Software, Formal analysis, Writing – original draft, Validation, Methodology, Data curation. **Kangwei Guo:** Visualization, Software, Formal analysis, Writing – original draft, Validation, Methodology, Data curation. **Wang Luo:** Methodology, Data curation, Supervision, Investigation. **Xiaojun Zeng:** Validation, Formal analysis, Conceptualization, Visualization, Supervision, Data curation. **Mingjun Lu:** Methodology, Data curation, Validation, Formal analysis. **Jinyu Lin:** Validation, Methodology, Visualization, Supervision. **Baihong Li:** Software, Investigation, Supervision, Methodology. **Yinling Qian:** Writing – review & editing, Software, Conceptualization, Supervision, Methodology. **Jian Yang:** Project administration, Funding acquisition, Writing – review & editing, Methodology, Conceptualization.

Funding

This study was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (Grant No. 2024ZD0525400); National Natural Science Foundation of China (Grant No. 82272132); Basic and Applied Basic Research Foundation of Guangdong Province (Grant No. 2024A1515012051 and 2023A1515110602); The Key Research and Development Plan Project of Guangzhou (Grant No. 2023B03J1246); China Postdoctoral Science Foundation (Grant No. 2022M721514 and 2024T170386).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2025.112415.

References

[1] E.A. Boonstra, M.T. de Boer, E. Sieders, P.M. Peeters, K.P. de Jong, M.J. Slooff, R. J. Porte, Risk factors for central bile duct injury complicating partial liver resection, Br. J. Surg. 99 (2) (2012) 256–262.
[2] K. Hayashi, Y. Abe, M. Shinoda, M. Kitago, H. Yagi, G. Oshima, S. Hori, T. Wakabayashi, Y. Kitagawa, Clinical impact of intraoperative bile leakage during laparoscopic liver resection, Surg. Endosc. 35 (8) (2021) 4134–4142.
[3] K.S. Chok, C.M. Lo, Prevention and management of biliary anastomotic stricture in right-lobe living-donor liver transplantation, J. Gastroenterol. Hepatol. 29 (10) (2014) 1756–1763.
[4] N. Kimura, A.L. Young, Y. Toyoki, J.I. Wyatt, G.J. Toogood, E. Hidalgo, K. R. Prasad, D. Kudo, K. Ishido, K. Hakamada, J.P.A. Lodge, Radical operation for hilar cholangiocarcinoma in comparable Eastern and Western centers: Outcome analysis and prognostic factors, Surgery 162 (3) (2017) 500–514.
[5] D. Cherqui, R. Ciria, C.H.D. Kwon, K.H. Kim, D. Broering, G. Wakabayashi, B. Samstein, R.I. Troisi, H.S. Han, F. Rotellar, O. Soubrane, J. Briceño, F. Alconchel, M.D. Ayllón, G. Berardi, F. Cauchy, I.G. Luque, S.K. Hong, Y.Y. Yoon, H. Egawa, J. Lerut, C.M. Lo, M. Rela, G. Sapisochin, K.S. Suh, Expert consensus guidelines on minimally invasive donor hepatectomy for living donor liver transplantation from innovation to implementation: a joint initiative from the international laparoscopic liver society (ILLS) and the asian-pacific hepato-pancreato-biliary association (A-PHPBA), Ann. Surg. 273 (1) (2021) 96–108.
[6] M. Bilgin, F. Shaikh, R.C. Semelka, S.S. Bilgin, N.C. Balci, A. Erdogan, Magnetic resonance imaging of gallbladder and biliary system, Top. Magn. Reson. Imaging 20 (1) (2009) 31–42.
[7] S. Morita, N. Saito, K. Suzuki, N. Mitsuhashi, Biliary anatomy on 3D MRCP: Comparison of volume-rendering and maximum-intensity-projection algorithms, J. Magn. Reson. Imaging 29 (3) (2009) 601–606.
[8] A. Becq, J. Szewczyk, G. Salin, M. Chartier, U. Chaput, R. Leenhardt, X. Dray, L. Arrive, M. Camus, ERCP 2.0: Biliary 3D-reconstruction in patients with malignant hilar stricture, Clin. Res. Hepatol. Gastroenterol. 47 (7) (2023) 102172.
[9] M.H. Goldfinger, G.R. Ridgway, C. Ferreira, C.R. Langford, L. Cheng, A. Kazimianec, A. Borghetto, T.G. Wright, G. Woodward, N. Hassanali, R. C. Nicholls, H. Simpson, T. Waddell, S. Vikal, M. Mavar, S. Rymell, I. Wigley, J. Jacobs, M. Kelly, R. Banerjee, J.M. Brady, Quantitative MRCP imaging: accuracy, repeatability, reproducibility, and cohort-derived normative ranges, J. Magn. Reson. Imaging 52 (3) (2020) 807–820.
[10] S. Bohlender, I. Oksuz, A. Mukhopadhyay, A survey on shape-constraint deep learning for medical image segmentation, IEEE Rev. Biomed. Eng. 16 (2023) 225–240.
[11] A. Pellat, M. Barat, R. Coriat, P. Soyer, A. Dohan, Artificial intelligence: a review of current applications in hepatocellular carcinoma imaging, Diagn. Interv. Imaging 104 (1) (2023) 24–36.
[12] J.G. Nam, J.M. Lee, H.J. Kang, S.M. Lee, E. Kim, J.M. Peeters, J.H. Yoon, GRASE Revisited: breath-hold three-dimensional (3D) magnetic resonance cholangiopancreatography using a Gradient and Spin Echo (GRASE) technique at 3T, Eur. Radiol. 28 (9) (2018) 3721–3728.
[13] F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-net: a self-configuring method for deep learning-based biomedical image segmentation, Nat. Methods 18 (2) (2021) 203–211.
[14] T. Nakamura, K. Tanaka, T. Kiuchi, M. Kasahara, F. Oike, M. Ueda, S. Kaihara, H. Egawa, I. Ozden, N. Kobayashi, S. Uemoto, Anatomical variations and surgical strategies in right lobe living donor liver transplantation: lessons from 120 cases, Transplantation 73 (12) (2002) 1896–1903.
[15] G.W. Ji, F.P. Zhu, K. Wang, C.Y. Jiao, Z.C. Shao, X.C. Li, Clinical implications of biliary confluence pattern for bismuth-corlette type IV hilar cholangiocarcinoma applied to hemihepatectomy, J. Gastrointest. Surg. 21 (4) (2017) 666–675.
[16] H.F. Al Janabi, A. Aydin, S. Palaneer, N. Macchione, A. Al-Jabir, M.S. Khan, P. Dasgupta, K. Ahmed, Effectiveness of the HoloLens mixed-reality headset in minimally invasive surgery: a simulation-based feasibility study, Surg. Endosc. 34 (3) (2020) 1143–1149.
[17] N. Zeng, H. Tao, C. Fang, Y. Fan, N. Xiang, J. Yang, W. Zhu, J. Liu, T. Guan, C. Fang, F. Xiang, Individualized preoperative planning using three-dimensional modeling for Bismuth and Corlette type III hilar cholangiocarcinoma, World J. Surg. Oncol. 14 (1) (2016) 44.
[18] T. Hyodo, S. Kumano, F. Kushihata, M. Okada, M. Hirata, T. Tsuda, Y. Takada, T. Mochizuki, T. Murakami, CT and MR cholangiography: advantages and pitfalls in perioperative evaluation of biliary tree, Br. J. Radiol. 85 (1015) (2012) 887–896.
[19] C. Fang, J. An, A. Bruno, X. Cai, J. Fan, J. Fujimoto, R. Golfieri, X. Hao, H. Jiang, L. R. Jiao, A.V. Kulkarni, H. Lang, C.R.A. Lesmana, Q. Li, L. Liu, Y. Liu, W. Lau, Q. Lu, K. Man, H. Maruyama, C. Mosconi, N. Örmeci, M. Pavlides, G. Rezende, J.H. Sohn, S. Treeprasertsuk, V. Vilgrain, H. Wen, S. Wen, X. Quan, R. Ximenes, Y. Yang, B. Zhang, W. Zhang, P. Zhang, S. Zhang, X. Qi, Consensus recommendations of three-dimensional visualization for diagnosis and management of liver diseases, Hepatol. Int. 14 (4) (2020) 437–453.
[20] J.A. Mazroua, Y.E. Almalki, M. Alaa, S.K. Alduraibi, M. Aboualkheir, A.S. Aldhilan, Z.A. Almushayti, S.A. Aly, M.A.A. Basha, Precision mapping of intrahepatic biliary anatomy and its anatomical variants having a normal liver using 2D and 3D MRCP, Diagnostics (basel) 13 (4) (2023).
[21] H. Yoen, J.M. Lee, S.M. Lee, H.J. Kang, J.S. Bae, E. Kim, J.M. Peeters, J.H. Yoon, Comparisons between image quality and diagnostic performance of 2D- and breath-hold 3D magnetic resonance cholangiopancreatography at 3T, Eur. Radiol. 31 (11) (2021) 8399–8407.
[22] J. Rhu, G.S. Choi, M.S. Kim, J.M. Kim, J.W. Joh, Image guidance using two-dimensional illustrations and three-dimensional modeling of donor anatomy during living donor hepatectomy, Clin. Transplant. 35 (1) (2021) e14164.
[23] N. Oh, J.H. Kim, J. Rhu, W.K. Jeong, G.S. Choi, J.M. Kim, J.W. Joh, 3D auto-segmentation of biliary structure of living liver donors using magnetic resonance cholangiopancreatography for enhanced preoperative planning, Int. J. Surg. 110 (4) (2024) 1975–1982.
[24] S.T. Fan, Liver functional reserve estimation: state of the art and relevance for local treatments: the Eastern perspective, J. Hepatobiliary Pancreat. Sci. 17 (4) (2010) 380–384.