

Artificial intelligence driven 3D reconstruction for enhanced lung surgery planning

Received: 17 July 2024

Accepted: 14 April 2025

Published online: 01 May 2025

 Check for updates

Xiuyuan Chen^{1,2,3,4,5,10}, Chenyang Dai^{6,10}, Muyun Peng^{7,10}, Dawei Wang^{8,10}, Xizhao Sui^{1,2,3,4,5}, Liang Duan⁶, Xiang Wang⁷, Xun Wang^{1,2,3,4,5}, Wenhan Weng^{1,2,3,4,5}, Shaodong Wang^{1,2,3,4,5}, Heng Zhao^{1,2,3,4,5}, Zhenfan Wang^{1,2,3,4,5}, Jiayi Geng^{1,2,3,4,5}, Chen Chen⁷, Yan Hu⁷, Qikang Hu⁷, Chao Jiang⁶, Hui Zheng⁶, Yi Bao⁶, Chao Sun⁹, Zhuoer Cui^{1,2,3,4,5}, Xiangyu Zeng^{1,2,3,4,5}, Huiming Han^{1,2,3,4,5}, Chen Xia⁸, Jinlong Liu⁸, Bing Yang⁸, Ji Qi⁸, Fanghang Ji⁸, Shaokang Wang⁸, Nan Hong⁹, Jun Wang^{1,2,3,4,5}, Kezhong Chen^{1,2,3,4,5}, Yuming Zhu^{6,11} , Fenglei Yu^{7,11}  & Fan Yang^{1,2,3,4,5,11} 

The increasing complexity of lung surgeries necessitates the need for enhanced imaging support to improve the precision and efficiency of preoperative planning. Despite the promise of 3D reconstruction, clinical adoption remains limited due to time constraints and insufficient validation. To address this, we evaluate an artificial intelligence-driven 3D reconstruction system for pulmonary vessels and bronchi in a retrospective, multi-center multi-reader multi-case study. Using a two-stage crossover design, ten thoracic surgeons assess 140 cases with and without the system's assistance. The system significantly improves the accuracy of anatomical variant identification by 8% ($p < 0.01$), reducing errors by 41%. Improvements in secondary endpoints are also observed. Operation procedure selection accuracy is improved by 8%, with a 35% decrease in errors. Preoperative planning time is decreased by 25%, and user satisfaction is high at 99%. These benefits are consistent across surgeons of varying experience. In conclusion, the artificial intelligence-driven 3D reconstruction system significantly improves the identification of anatomical variants, addressing a critical need in preoperative planning for thoracic surgery.

Surgery is the optimal treatment for most pulmonary lesions, particularly lung cancer, the leading cause of cancer-related mortality worldwide. With an estimated annual global demand exceeding one million operations^{1,2}, the increasing prevalence of early-stage lung cancer^{3,4} and a growing shortage of thoracic surgeons⁵ necessitates innovative solutions to enhance surgical efficiency and safety. Precise preoperative planning is the foundation to achieving this objective⁶, with accurate identification of anatomical variants and appropriate selection of operation procedures being critical steps.

Two-dimensional (2D) computed tomography (CT) remains the primary tool for preoperative planning in lung surgery. However, the inherent limitations of 2D imaging hinder intuitive visualization of complex anatomical structures, and potentially leading to misidentification, particularly within the distal pulmonary vasculature where anatomical variations are increasingly complex⁷. Three-dimensional (3D) reconstructions offer improved spatial understanding, facilitating more accurate surgical planning and execution^{7,8}. Similar benefits have been observed in other surgical specialties,

A full list of affiliations appears at the end of the paper.  e-mail: drzhuyuming@163.com; yufenglei@csu.edu.cn; dr.yangf@gmail.com

including urological⁹, oesophagogastric¹⁰, head and neck¹¹, and pancreatic surgeries¹². Despite these advantages, the routine adoption of 3D reconstructions remains constrained by the time-intensive nature of manual image segmentation¹³, resulting in utilization rates below 25% in major operations despite recognized benefits¹¹.

The integration of artificial intelligence (AI) algorithms represents a critical advancement in 3D reconstruction, providing time-efficient and accurate models that match or exceed the performance of manual methods^{7,8,13}. This has led to investigations into the clinical impact of AI-driven 3D models on both preoperative and perioperative outcomes. Our prior pilot study indicated that with AI driven 3D reconstruction, surgeons can achieve an accuracy of 85% in identifying anatomical variants⁷, compared to 78% using 2D CT¹³. Wang et al.¹⁴ have reported that 3D reconstruction may reduce operation time by 12.4%, decrease stapler reload by 13.4%, and lower air leakage ratio by 61.5%. Similarly, Li et al.⁸ have reported a 17.2% reduction in operation time with AI driven 3D reconstruction. Contradictory evidence was also reported. A randomized controlled trial (RCT) found no significant difference in operative times with or without AI driven 3D reconstruction¹⁵, raising questions about the magnitude and consistency of perioperative benefits. This discrepancy, potentially attributable to limited statistical power in the RCT, underscores the need to rigorously evaluate the *direct* impact of AI driven 3D reconstruction on preoperative planning before extrapolating to downstream perioperative outcomes.

To address this issue, we conduct a retrospective multi-center, multi-reader, multi-case (MRMC) study. In this study, we show the effectiveness of an AI-driven 3D reconstruction system (referred to as the AI-3D system) in improving the accuracy of anatomical structure identification and operation procedure selection during preoperative planning for anatomical lobectomy and segmentectomy.

Result

Baseline characteristics of the MRMC study

A total of 450 patients from three independent medical centers were consecutively enrolled according to inclusion and exclusion criteria, spanning the period from July 2021 to January 2022. From this cohort, 140 cases were randomly selected. The median age of the patient cohort was 58 years, with an interquartile range of 48.75 to 65 years. Among the participants, there were 76 women (54.3%) and 64 men (45.7%). Regarding the surgical procedures, 62 patients underwent lobectomy, while 78 patients underwent segmentectomy. The nodules were distributed across all lobes, with 49 (35%) located in the right upper lobe, 4 (2.9%) in the right middle lobe, 24 (17.1%) in the right lower lobe, 39 (27.9%) in the left upper lobe, and 24 (17.1%) in the left lower lobe. The most frequently observed histological type of lesion was invasive adenocarcinoma, accounting for 83 cases (59.3%). A comprehensive overview of the baseline characteristics is shown in Table 1.

Ten thoracic surgeons from three medical centers participated as readers. The age of the readers spanned from 34 to 45, and the years of practicing ranged from 6 to 19 years (Table S1). A schematic of the study is shown in Fig. 1.

Anatomical variant identification

In the primary analysis, AI-3D assistance exhibited a superior case-wise median accuracy of 0.87 compared to 0.78 without AI-3D in anatomical variant identification ($p < 0.01$, Fig. 2A). This improvement corresponded to a 41% reduction in identification error (RR = 0.59, 95% CI = 0.56 – 0.63). Consistent improvement was observed across 10 readers (Fig. 2B), 5 lobes (Fig. S2A) and 90% (35 / 39) anatomical structures (Fig. 2C, and Supplementary figs. S2B–D). An illustrative case demonstrating this improvement is shown in Fig. S2E and its result in Table S2. These results indicate that AI-3D assistance significantly enhances surgeons' ability to accurately identify anatomical variants during preoperative planning.

Table 1 | Characteristics of the enrolled patients

| Variable | Total | Center 1 | Center 2 | Center 3 |
|-------------------------------------|-------------|-------------|-------------|---------------|
| Number of cases, n | 140 | 60 | 40 | 40 |
| Age, median (IQR), years | 58 (49, 65) | 60 (53, 65) | 56 (47, 65) | 52.5 (46, 65) |
| Sex, n (%) | | | | |
| Female | 76 (54.3%) | 26 (43.3%) | 28 (70.0%) | 22 (55.0%) |
| Male | 64 (45.7%) | 34 (56.7%) | 12 (30.0%) | 18 (45.0%) |
| Tumor location | | | | |
| Right upper lobe | 49 (35.0%) | 20 (33.3%) | 14 (35.0%) | 15 (37.5%) |
| Right middle lobe | 4 (2.9%) | 1 (1.7%) | 1 (2.5%) | 2 (5.0%) |
| Right lower lobe | 24 (17.1%) | 15 (25.0%) | 5 (12.5%) | 4 (10.0%) |
| Left upper lobe | 39 (27.9%) | 14 (23.3%) | 11 (27.5%) | 14 (35.0%) |
| Left lower lobe | 24 (17.1%) | 10 (16.7%) | 9 (22.5%) | 5 (12.5%) |
| Surgery type, n (%) | | | | |
| Segmentectomy | 78 (55.7%) | 34 (56.7%) | 18 (45.0%) | 26 (65.0%) |
| Lobectomy | 62 (44.3%) | 26 (43.3%) | 22 (55.0%) | 14 (35.0%) |
| Histology, n (%) | | | | |
| ADC | 83 (59.3%) | 46 (76.7%) | 18 (45.0%) | 19 (47.5%) |
| MIA | 28 (20.0%) | 12 (20.0%) | 1 (2.5%) | 15 (37.5%) |
| AIS | 12 (8.6%) | 1 (1.7%) | 10 (25.0%) | 1 (2.5%) |
| SCC | 7 (5.0%) | 1 (1.7%) | 5 (12.5%) | 1 (2.5%) |
| AAH | 2 (1.4%) | 0 | 1 (2.5%) | 1 (2.5%) |
| LCNEC | 2 (1.4%) | 0 | 1 (2.5%) | 1 (2.5%) |
| Metastatic | 1 (0.7%) | 0 | 1 (2.5%) | 0 |
| Benign | 4 (2.9%) | 0 | 3 (7.5%) | 1 (2.5%) |
| Others | 1 (0.7%) | 0 | 0 | 1 (2.5%) |
| CT instrument manufacturer | | | | |
| GE | 45 (32.1%) | 12 (20.0%) | 33 (82.5%) | 0 |
| Philips | 35 (25.0%) | 28 (46.7%) | 7 (17.5%) | 0 |
| Siemens | 52 (37.1%) | 19 (31.7%) | 0 | 33 (82.5%) |
| UIH | 8 (5.7%) | 1 (1.7%) | 0 | 7 (17.5%) |
| Chest CT slice thickness, mm | | | | |
| 1 | 92 (65.7%) | 47 (78.3%) | 7 (17.5%) | 38 (95.0%) |
| 1.25 | 13 (9.3%) | 13 (21.7%) | 0 | 0 |
| 0.7 | 2 (1.4%) | 0 | 0 | 2 (5.0%) |
| 0.625 | 33 (23.6%) | 0 | 33 (82.5%) | 0 |

SCC Squamous cell carcinoma.

AAH Atypical adenomatous hyperplasia.

AIS Adenocarcinoma in situ.

MIA Microinvasive adenocarcinoma.

ADC Invasive adenocarcinoma.

LCNEC Large cell neuroendocrine carcinoma.

Source data are provided as a Source Data file.

A positive correlation was found between identification accuracy and variant prevalence ($R^2 = 0.68$, 95% CI = 0.59 – 0.76, Fig. 2D). Interestingly, a trend toward larger improvement by AI-3D system was observed in variants with lower than with higher prevalence (Fig. 2E), indicating a potential benefit for AI-3D in identifying less common variants (Table S3, S4).

Identification errors in anatomical variants can arise from two reasons: misidentification of the variant and missing value. The latter occurs due to incorrect selection of the operation procedure, which can introduce potential bias into the analysis. To address this issue, we excluded cases with unanswered structure and re-evaluated the impact of AI-3D assistance on anatomical variant identification. Notably, the beneficial effects of AI remained evident after this exclusion (AI vs Ctrl: 0.86 vs 0.79, RR = 0.64, 95% CI = 0.57 – 0.82, Fig. S2F, S2G), demonstrating that the improvements in accuracy are robust and not merely artifacts of procedural errors.

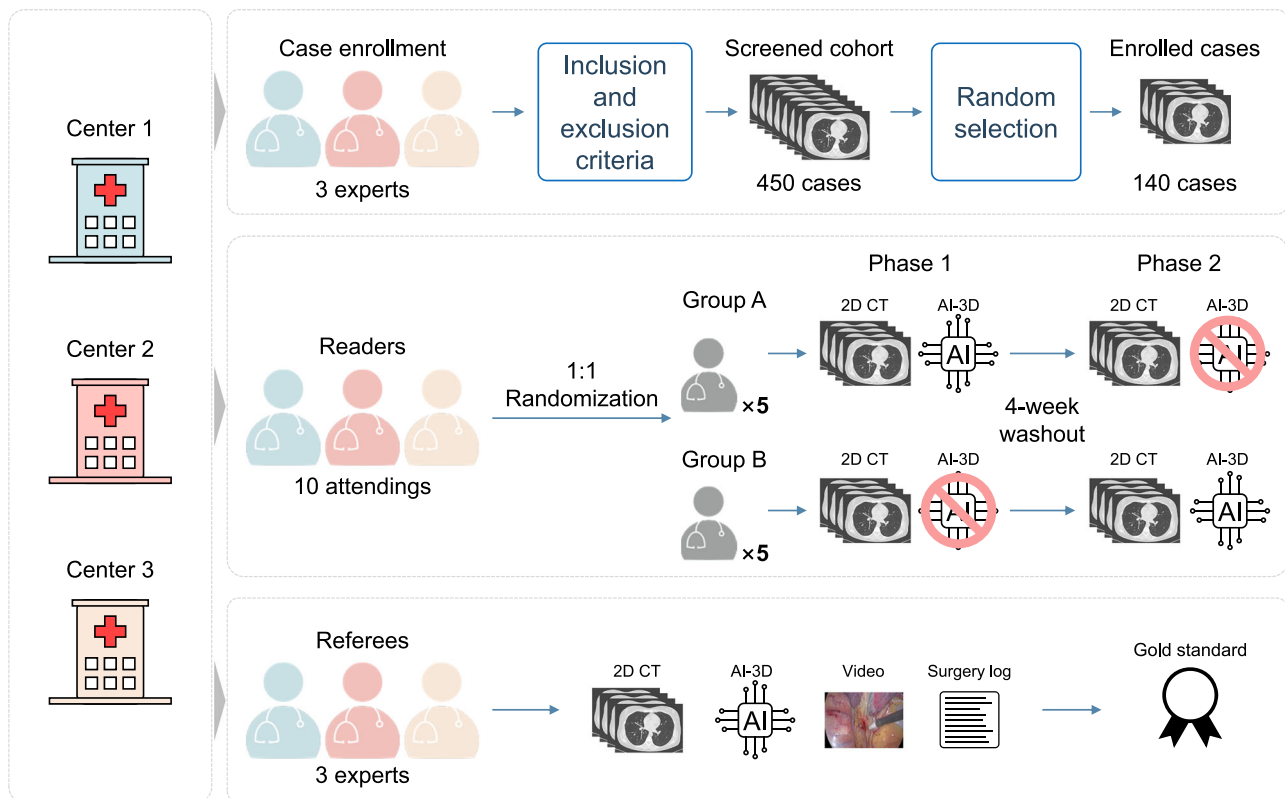


Fig. 1 | Schematic illustration of the multi-reader multi-cases (MRMC) study. A total of 16 thoracic surgeons from three top-tier hospitals in China participated in this study. Three surgeons collected 450 patients and enrolled in the study according to inclusion and exclusion criteria. A total of 140 cases were randomly selected from eligible cohorts. Three expert surgeons from three centers established the golden standard based on collected surgical logs, videos, CT scans, and

manually constructed 3D reconstructions. The other 10 surgeons were divided into two groups randomly and took part in the MRMC study in which two rounds of fully crossed pre-operative planning simulations were conducted with an interval of 4 weeks, either with or without the aid of AI-derived 3D reconstructions of pulmonary vessels.

Operation procedure selection

The accuracy for operation procedure selection was improved from 0.77 to 0.85 with AI-3D assistance (estimated improvement 0.08, 95% CI = 0.04 – 0.12, Fig. 3A, B). This improvement corresponded to a 35% reduction in error (RR = 0.65, 95% CI = 0.54 – 0.77). In a binary comparison of the selection between lobectomy and segmentectomy, a 0.04 (95% CI = 0.01 – 0.07) increase in the accuracy was noted under AI-3D assistance (Fig. 3C). The improvement was more pronounced in lobectomy compared to segmentectomy cases (Fig. S3A) as more correction rather than misleading by AI-3D assistance was seen in lobectomy cases (Fig. S3B). An illustrative case is shown in Fig. S3C and its result in Table S5.

A comprehensive error analysis revealed that AI-3D assistance improved the accuracy of operation procedure selection in favor of certain error types (Fig. 3D, and Supplementary figs. S3D, S3E). Mistaken resection, defined as resecting the wrong lesion, was reduced by 73% (RR = 0.27, 95% CI = 0.16 – 0.45, Fig. 3E). Insufficient resection, characterized by inadequate resection margin, was decreased by 51% (RR = 0.49, 95% CI = 0.38 – 0.70, Fig. 3E). On the contrary, the reduction in excessive resection, which involves removing more lung than necessary, was minimal at 2% (RR = 0.98, 95% CI = 0.77 – 1.26, Fig. 3E). These findings highlight the primary role of AI in operation procedure selection: it effectively pinpoints the target lesion and minimizes the risk of insufficient resection due to misinterpretation of resection margins.

Time consumption, confidence, and interobserver agreement

AI-3D assistance reduced the median time consumption for surgery planning by 63 seconds (95% CI = 42 – 78), representing a 25% decrease

(RR = 0.75, 95% CI = 0.69 – 0.80, Figure 4A, B). Notably, cases that initially required longer planning times were associated with lower accuracy, but these cases showed more substantial improvements with AI-3D assistance, indicating that the AI system is particularly beneficial in more challenging scenarios. (Fig. S4). The inference time for the 3D reconstruction itself was 233.76 ± 75.08 seconds. While a direct comparison is beyond the scope of this study, prior experience suggests manual segmentation could take up to 30 minutes.

In the post-hoc analysis of confidence, cases were categorized as confident (confidence score = 100) and non-confident (confidence score < 100), allowing for a binary analysis. The use of AI notably increased the proportion of confident cases in both anatomical variant identification (RR = 1.31, 95% CI = 1.29 – 1.34, Fig. 4C) and operation procedure selection (0.67 vs 0.57, RR = 1.17, 95% CI = 1.10 – 1.25, Fig. 4C). Higher confidence was correlated with higher accuracy in both anatomical variant identification and operation procedure selection (Fig. 4D, E).

Agreement among readers was significantly improved by AI-3D assistance in both anatomical variant identification ($\kappa = 0.43$ vs 0.33) and operation procedure planning ($\kappa = 0.76$ vs 0.70) (Table 2).

Satisfaction with the AI-3D system

As for the independent performance of the model, the overall satisfaction was 99%. Among the 1% (14/1400) cases that were reported with unsatisfactory reconstruction performance, 43% (6/14) were structure masked by the pulmonary lesion, 29% (4/14) were unsatisfied distal branches, 14% (2/14) were classification errors (pulmonary veins misidentified as arteries), 14% (2/14) were unclear demonstration without detailed description. Unsatisfactory reconstruction posed minimal influence on the accuracy (0.82) of identifying anatomical variants.

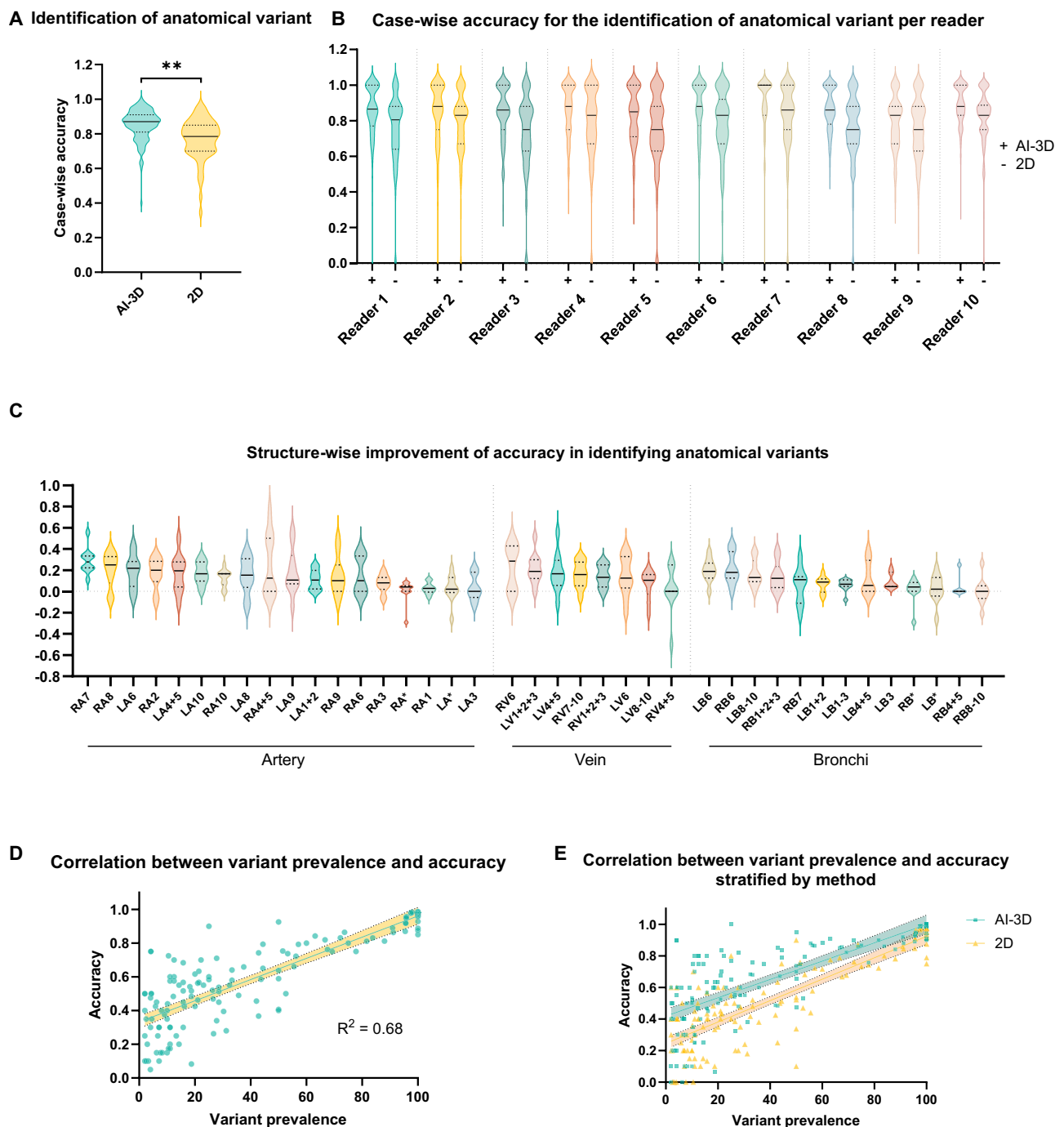


Fig. 2 | The anatomical variant identification analysis. **A** Significant improvement of the overall case-wise accuracy of anatomical variant identification. (** stands for statistical significance $p < 0.01$. Two-sided Mann-Whitney test was used). **B** Similar degree of improvement of the case-wise accuracy of anatomical variant identification in each reader. **C** The improvement in accuracy among 35 / 39 anatomical structures. **D** Positive correlation existed between variant prevalence and identification accuracy.

(The error bar stands for the 95% CI of the regression line). **E** Higher accuracy in anatomical variant identification was observed with AI-3D assistance comparing to control method across different variant prevalence. The difference tended to be larger in lower prevalence variants. (The error bar stands for the 95% CI of the regression line.) Source data are provided as a Source Data file.

The high satisfaction rate underscores the usability and reliability of the AI-3D system in clinical practice.

Discussion

In this study, we highlighted the pronounced positive impact of AI-3D assistance in improving anatomical variant identification and operation procedure selection across medical centers. Our results demonstrated accuracy improvements averaged around 10%, leading to a

substantial reduction in errors ranging from 36% to 52%. Such enhancements are crucial for surgeons, as even minor errors can have severe repercussions for patient outcomes. Therefore, the reduction in errors achieved by the AI-3D system is both statistically significant and clinically meaningful.

AI has emerged as a crucial factor in augmenting the accessibility of 3D reconstruction in medical imaging. Traditional computer vision methods were hindered by several drawbacks: they are time-intensive,

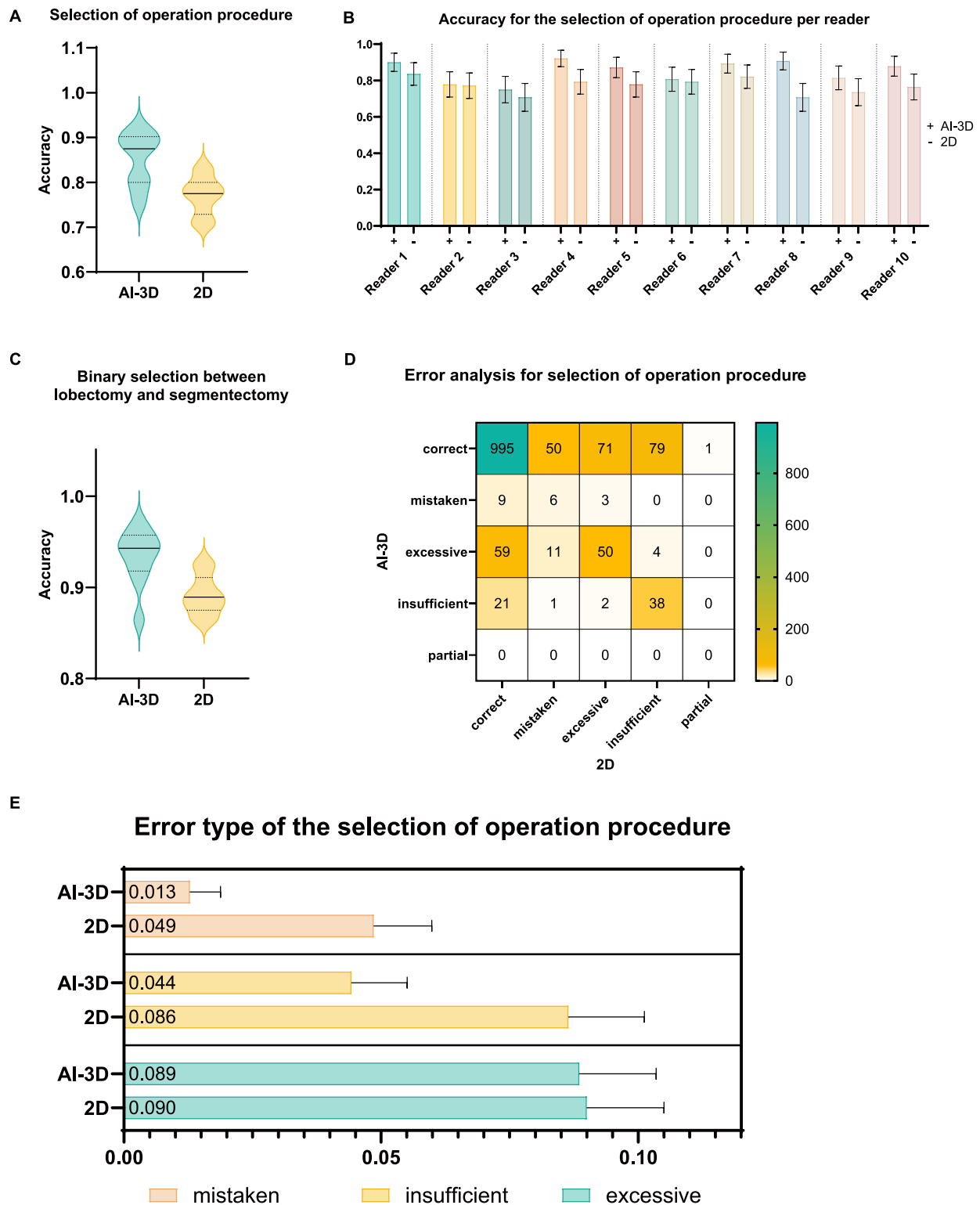


Fig. 3 | The operation procedure selection analysis. **A** The accuracy of operation procedure selection was improved by AI-3D assistance. **B** The accuracy of operation procedure selection of each reader was improved by AI-3D assistance. ($n = 140$, the error bar stands for the 95% CI). **C** Binary selection between lobectomy and segmentectomy was improved by AI-3D assistance. **D** The heatmap demonstration of

error types of resection extent determination with or without AI-3D. **E** Insufficient and mistaken resection were profoundly reduced with AI-3D assistance. ($n = 1400$, the error bar stands for the upper side of the 95% CI, the lower side is hidden to avoid overlaying numbers presented in the chart). Source data are provided as a Source Data file.

often requiring 0.5 to 2 hours for reconstruction^{16,17}, which was impractical in busy clinical environments; the reconstruction quality was often impaired by artifacts¹⁸ such as staircase effects or blockiness that required manual correction; accuracy was limited, especially in

non-contrast CT scans, as these methods depend on the contrast among arteries, veins and soft tissue. The integration of AI has addressed these issues by profoundly enhancing efficiency, decreasing artifacts, and improving accuracy in non-contrast CT^{16,19,20}. In our

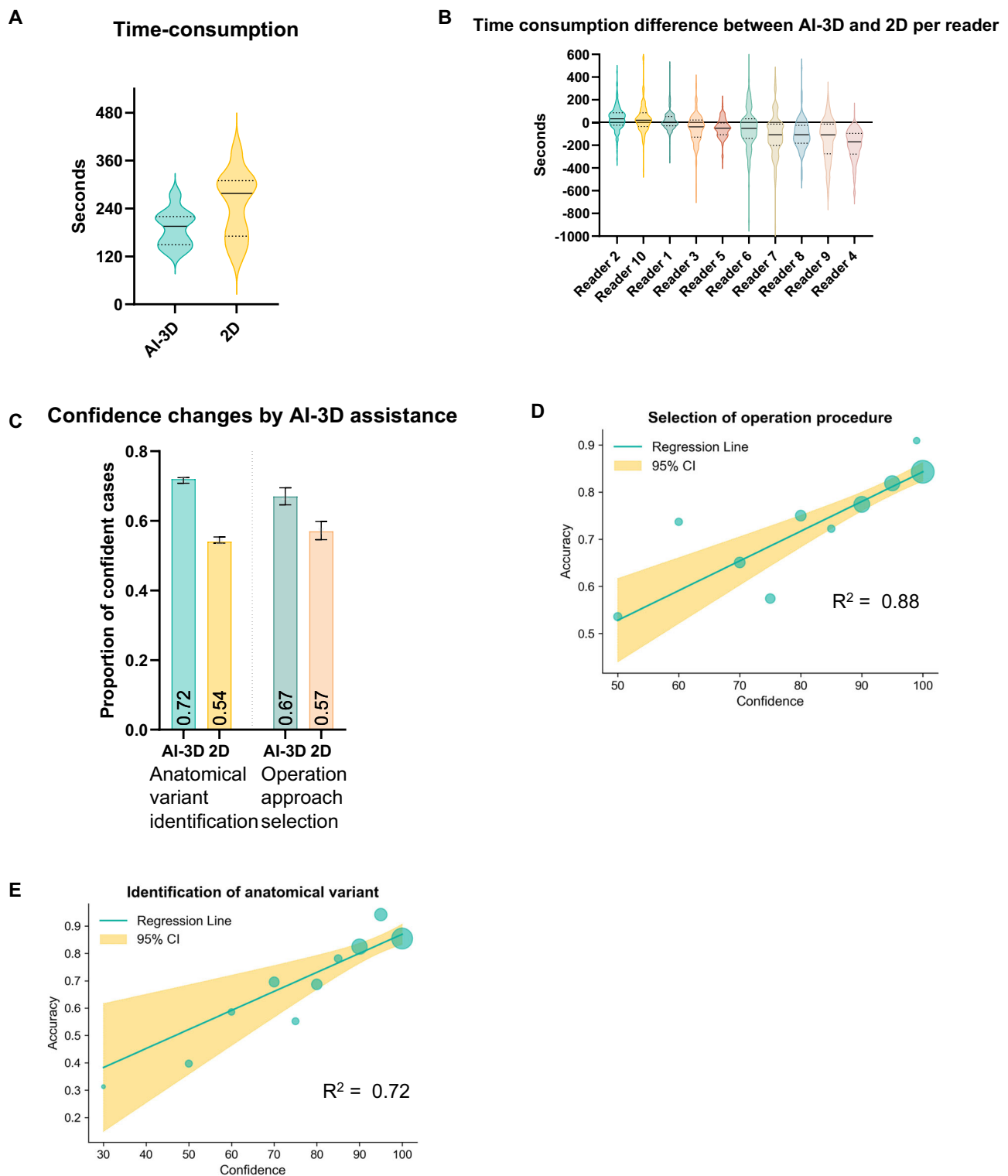


Fig. 4 | The time consumption and confidence analysis. **A** The overall time consumption was decreased by AI-3D assistance. **B** Time consumption difference between AI-3D and 2D approach among 10 readers. **C** The confidence of operation procedure selection was improved by AI-3D assistance. (n for anatomical variant identification are 11,903 with AI-3D and 11471 with 2D; n for operation approach selection is 1382 with or without AI-3D; the error bar stands for the 95% CI).

D Positive correlation was observed between operation procedure selection and readers' confidence level. (The error bar stands for the 95% CI of the regression line). **E** Positive correlation was observed between anatomical variant identification and readers' confidence level. (The error bar stands for the 95% CI of the regression line). Source data are provided as a Source Data file.

study, the time required for the reconstruction process was reduced by 20-fold, and image interpretation time was decreased by 25%. Compared to traditional manual reconstruction, these enhancements saved more than 30 minutes in the planning process. Furthermore, the

reconstruction quality and accuracy were satisfactory in 99% of cases. These findings suggest that the integration of AI algorithm has strengthened 3D reconstruction, making it an accessible and reliable auxiliary tool for thoracic surgery planning.

Table 2 | Interobserver agreement

| | AI-3D (95% CI) | 2D (95% CI) | P value |
|---------------------------|-------------------------|-------------------------|---------|
| Operation procedure | 0.762 (0.722, 0.802) | 0.695 (0.645, 0.740) | < 0.01 |
| Anatomical variant | | | |
| Overall | 0.434 (0.413, 0.457) | 0.329 (0.309, 0.350) | < 0.01 |
| High prevalence variant | 0.330 (0.286, 0.372) | 0.293 (0.265, 0.321) | < 0.01 |
| Low prevalence variant | 0.317 (0.288, 0.346) | 0.212 (0.185, 0.240) | < 0.01 |

Source data are provided as a Source Data file.

The indication of using AI-3D should be considered from both patients’ and surgeons’ perspectives. While our results showed significant benefits across all patients, certain subpopulations may particularly benefit from AI-3D assistance. Patients with low prevalence anatomical variants, cases that typically require longer planning times, and scenarios where surgeons have lower confidence levels are more prone to inaccuracies. These factors may serve as practical indicators for utilizing AI-3D reconstruction algorithms. However, given the potentially devastating consequences of misidentifying anatomical structures, such as the mis-ligation of arteries or massive bleeding, we advocate using the AI-3D system in both lobectomy and segmentectomy cases, especially in more complex and unfamiliar procedures, such as complex segmentectomies.

Our study showed that AI-3D assistance is beneficial for surgeons across all experience levels. Although some studies have demonstrated that less experienced clinicians (e.g., trainees), benefit more from AI assistance than their more experienced peers^{21–23}, our results showed a significant impact on surgeons who have been practicing for 6 – 19 years, ranging from junior to senior attendings alike. Notably, interobserver agreement significantly improved with AI-3D assistance, indicating that the system helps decrease errors and unify surgical decisions regardless of experience level. Furthermore, the improvement observed in our study quantitatively aligns with findings from previous publications^{8,20}, suggesting that this approach may be widely applicable among various surgeons.

Although no unforeseen complications introduced by human-AI interaction were observed, intrinsic risks in applying AI in medicine should not be neglected. Several studies have underscored an increase in sensitivity with a decrease in specificity^{21,24–27}. This was particularly evident in the early AI application in thoracic surgery such as pulmonary nodule detection. These algorithms struggled to minimize false positives. For convenience, rule-based constraints were usually implemented, sacrificing the flexibility and generalizability of the algorithm¹⁶. In our study, the risk of misleading in selection of operation procedures, especially for potential segmentectomy candidates, should be noted. While regarding anatomical variant identification, the AI-3D system outputs the anatomical structure instead of a binary conclusion. The output serves as an intermediate step in the surgery planning process and is subsequently interpreted by the surgeon, and thus shows minimized misleading risk. This success highlighted the advantage of the human-in-the-loop paradigm where AI works collaboratively with human expertise, rather than replacing it²⁸.

This study has several limitations. First, while the identification of anatomical variants is crucial for surgical planning, it does not directly translate to improved surgical outcomes, such as reduced intraoperative bleeding, shorter operative times, or fewer complications. Previous studies have yielded mixed results regarding the impact of AI-driven 3D reconstruction on surgical outcomes^{8,15}. For instance, while Li et al.⁸ reported improvements in operative times, the only randomized controlled trial (RCT) to date found no significant difference in operative times with or without AI-3D assistance¹⁵. This conclusion,

however, may be due to an underpowered study design. The hypothesized 14% decrease in operative time is larger than the degree of the benefit we observed in preoperative planning (8% increase in identification accuracy). Additionally, the predominance of simple segmentectomy cases in the RCT likely diluted the potential effects of AI-assisted preoperative planning. Furthermore, most errors in identifying anatomical variants during planning are corrected intraoperatively, which may further obscure the impact of AI-3D assistance on surgical outcomes. Nevertheless, it is reasonable to infer that improved preoperative understanding of anatomical structures could improve surgical outcomes. In cases involving interlobular arteries or veins, prior knowledge could prevent bleeding caused by blunt dissection of the interlobular fissure; in segmentectomy cases, the misidentification of segmental pulmonary arteries and veins may cause mis-ligation of these structures. A better understanding of anatomical structures may reduce the need for intraoperative observation and potentially shorten operative times, as was noted in Li et al.’s study⁸. Future research, including RCTs for a certain subset of patients, is required to establish a more definitive relationship between AI-3D assistance and patient outcomes.

Second, the MRMC study underrepresented rare anatomical variants (e.g., independent upper pulmonary vein, bronchus suis, pulmonary artery sling), and the algorithm’s and surgeons’ performance in such cases remains to be evaluated. A dedicated database of these rare variants is currently being compiled for future validation.

Third, the selection of surgical procedure was based on a resection margin no less than the diameter of the nodule on chest CT. However, in real-world clinical practice, the ideal resection margin is more closely related to the pathological margin and the local recurrence rate. These important considerations fall beyond the scope of this study and warrant further investigation in future research.

Fourth, although exploratory analyzes suggested a greater benefit for less common variants and less confident cases, these analyzes were not pre-specified. Furthermore, the current sample size may be insufficient to definitively establish the presence of these beneficial subgroups. While these trends offer valuable potential directions for future research, they require confirmation in a larger, pre-specified study.

Finally, the observed improvements in our study were less pronounced than clinically anticipated. This may be due to the inherent limitations of MRMC design, which may not fully capture the complexities of real-world clinical workflows, including data interpretation sequence, workload, teamwork dynamics, and consequences of misinterpretation. These factors can introduce bias in assessing the perceived value of AI-3D reconstruction^{29,30}. Nevertheless, the MRMC design offers a valuable balance between controlling for bias and reflecting real-world variability, enabling a rigorous evaluation of the AI-3D system’s clinical value³⁰.

In conclusion, the AI-3D system enhances the accuracy of anatomical variant identification during preoperative planning for lung surgery. This improvement, coupled with increased accuracy in operation procedure selection and a reduction in planning time, benefits surgeons across a range of experience levels. While these findings suggest the potential for improved surgical outcomes and expanded surgical capacity, further research, including prospective trials incorporating perioperative patient outcomes is necessary to definitively establish the clinical value of the AI-3D system.

Methods

Ethics statement

The study was approved by the Institutional Review Boards of Peking University People’s Hospital (2021PHA038-002), Shanghai Pulmonary Hospital (21Q035XW-1), and the Second Xiangya Hospital of Central South University (2021-K380). A waiver of consent was granted due to the retrospective nature and minimal risk of the study. All data were anonymized to protect patients’ privacy.

Study design and participants

The study was approved by the Institutional Review Boards of Peking University People's Hospital (2021PHA038-002), Shanghai Pulmonary Hospital (21Q035XW-1), and the Second Xiangya Hospital of Central South University (2021-K380).

This retrospective, multi-center, MRMC study (Fig. 1) was designed to evaluate the effectiveness of an automatic surgery planning software for chest computed tomography (CT) images (the AI-3D system). The primary objective was to compare the accuracy of thoracic surgeons in identifying anatomical structures during preoperative planning for anatomical lobectomy and segmentectomy, with and without the assistance of the AI-3D system.

Patients were retrospectively enrolled from three centers: Peking University People's Hospital, Shanghai Pulmonary Hospital, and the Second Xiangya Hospital of Central South University, between July 2021 and January 2022. Eligible patients were adults aged 18 years or older who had undergone anatomical lobectomy or segmentectomy and had high-quality chest CT images in Digital Imaging and Communications in Medicine (DICOM) format with a slice thickness of ≤ 2 mm and no interval reconstruction. Exclusion criteria included patients with prior lung surgery or chest trauma, poor-quality imaging due to non-standard scanning or motion artifacts, trans-lobar invasion of target nodules, or incomplete clinical data (Supplementary note 1).

A total of 450 patients from 3 centers were retrospectively collected based on inclusion and exclusion criteria by 3 thoracic surgeons. From this pool, 140 patients were randomly selected to constitute the reader study dataset.

Ten board-certified thoracic surgeons participated as readers in the study. All readers held at least the title of attending physician, had a minimum of one year of experience in thoracoscopic surgery, possessed medical practitioner qualification certificates, and had Good Clinical Practice certification.

An expert panel of three senior thoracic surgeons, independent of the readers, established the gold standard by reviewing 2D CT, manual 3D reconstruction, surgical videos and logs. The surgical margin was suggested to be no less than the diameter of the nodule on chest CT. The panel consisted of thoracic surgeons with associate senior clinical titles or above. Disagreements were resolved through arbitration by a third expert to establish the gold standard.

AI-3D system for thoracic surgery planning

The AI-3D system (InferOperate Thorax) was developed using deep learning algorithms (Supplementary note 2, Fig. S1). For bronchial segmentation, a patch-based three-dimensional U-Net architecture was employed. Lung vessel segmentation combined 2.5-dimensional and three-dimensional models to optimize performance and computational efficiency. Specifically, the 2.5D model segmented pulmonary blood vessels within the lung parenchyma, while the 3D network focused on mediastinal vessels. A region-growing algorithm extended the 3D segmentation results into the peripheral lungs (Fig. S1). The segmented bronchi and vessels were converted into a triangular patch using MarchingCube, and RayCasting with DepthPeeling was used for efficient 3D rendering to assist surgical planning. A lesion detection module previously developed was embedded in the system.

The algorithmic performance of the system was detailed in Supplementary note 2.

Randomization and procedures

The study employed a two-stage, fully crossed, crossover design to reduce bias. Readers were randomly assigned to two groups using a computer-generated randomization schedule. In the first stage, Group A performed preoperative planning for all cases with AI-3D assistance, while Group B planned without AI-3D assistance. After a washout period of at least 28 days to mitigate memory bias, the 2 groups crossed over in the second stage: Group A planned without AI-3D

assistance, and Group B with AI-3D assistance. The sequence of case presentation was randomized in each stage and the surgery performed in the real world was blinded to further minimize potential biases.

Prior to the study, readers received standardized training on the use of the AI-3D system and the assessment procedures. Training included operating the software, interpreting AI-3D reconstructions, and completing the standardized surgical planning forms.

Readers assessed each case using a standardized surgical planning form focused on identifying anatomical variants and selecting the operation procedure (Supplementary note 3). The time taken for each planning session was recorded in seconds using a screen timing device. For each question, readers assigned a confidence score based on their certainty in their planning decisions, with higher scores reflecting greater confidence. After each use of the AI-3D system, readers completed a questionnaire to evaluate user satisfaction on a 5-point Likert scale.

End points

The primary endpoint was the case-wise accuracy of anatomical structure identification. Accuracy per case was calculated as the number of anatomical structures correctly identified by the reader divided by the total number of anatomical structures determined by the expert panel.

Secondary endpoints included 1) Time efficiency: The time required for anatomical structure identification during preoperative planning, measured from the start of planning to the completion of the anatomical structure identification form; 2) Surgical procedure selection accuracy: The consistency rate between the readers' surgical procedure determinations and those of the expert panel; 3) User satisfaction: Surgeons' satisfaction levels with the AI-3D system, assessed using a Likert scale ranging from 1 (very dissatisfied) to 5 (very satisfied).

Statistical analysis

Based on a pilot study involving four thoracic surgeons and 20 patients, which showed an effect size of 0.13 in the accuracy difference of identifying anatomical structures with and without the AI-3D system, a conservative effect size of 0.7 within each of the ten participating surgeons was assumed. Using these parameters, it was determined that at least 112 chest CT scans were required to achieve 80% power at a two-sided significance level of 5%. To account for potential dropouts and missing data, 140 cases were included.

The Dorfman-Berbaum-Metz-Hillis (DBMH) method was used to analyze the effectiveness of the AI-3D system in the multi-center MRMC study. The null hypothesis was that the readers' average identification accuracy with AI-3D assistance was less than or equal to that without AI-3D assistance. Continuous data were compared using t-tests for normal distributions (presented as mean \pm standard deviation) and Mann-Whitney tests for skewed distributions (presented as median [P25, P75]). Simple linear regression was used for regression tasks, with Weighted Least Squares (WLS) regression employed when data count disparities were too large. Fleiss' Kappa statistic was used to assess interobserver agreement among multiple readers. $P < 0.05$ was considered statistically significant. Data was analyzed using SPSS 23.0, SAS 9.4, Python, and R studios. An independent review committee assessed the primary and secondary endpoints of the MRMC in parallel with the investigator.

Missing data were addressed as follows: In the primary endpoint analysis assessing the accuracy of anatomical variant identification, 918 of 24,400 (3.8%) instances had missing identifications. These were considered misidentifications, reflecting a worst-case scenario. An exploratory analysis excluding these cases was also conducted to assess the robustness of the primary endpoint. Additional missing data points were excluded from their respective analyzes, including time consumption data (24/2,800, 0.9%), confidence ratings for operation

procedure selection (18/2,800, 0.6%) and for variant identification (1026/24,400, 4.2%).

Study registration

This study, conducted for National Medical Products Administration product registration, was not registered in a public database to protect trade secrets. The clinical study complied with Good Clinical Practice provisions and completed the premarket clinical study in three centers. The product registration acceptance number is CQZ2301043. Data entry and management were completed by the Statistics Teaching and Research Department of Capital Medical University, while data analysis and statistical analysis reports were completed by the Biostatistics Department of the School of Public Health of Peking University, both independent third parties.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The source imaging data obtained from three tertiary hospitals are restricted from public access due to privacy protection and upcoming research. Requests for data sharing can be submitted to F.Y. for collaborative purposes. Source data are provided with this paper.

Code availability

InferOperate Thorax is an intellectual property of Infervision Medical Technology Co., Ltd and thus restrictions apply to the source codes. As proprietary software, it is not available under an open-source license. An online version of InferOperate Thorax is available at <https://inferx.infervision.eu/>. The user guide of the system is provided as an independent Supplementary Movie 1. A total of 10 free trial case limit was set. Request for additional software trial access can be submitted to X.C at dr_chenxy@pku.edu.cn.

References

- Potter, A. L. P. T. et al. Assessing the number of annual lung cancer resections performed in the United States. *Shanghai Chest* **7**, 29 (2023).
- Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA: a Cancer J. Clinicians* **73**, 17–48 (2023).
- Ganti, A. K., Klein, A. B., Cotalra, I., Seal, B. & Chou, E. Update of incidence, prevalence, survival, and initial treatment in patients with non-small cell lung cancer in the US. *JAMA Oncol.* **7**, 1824–1832 (2021).
- Xia, C. et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin. Med J. (Engl.)* **135**, 584–590 (2022).
- Koch, R., Felsted, A. M., Virk, S., Roy, N. & Jayaraman, S. Surgical capacity building in low- and middle-income countries: lessons for thoracic surgery. *Thorac. Surg. Clin.* **32**, 269–278 (2022).
- Wasik, J., Tubbs, R. S., Zielinska, N., Karauda, P. & Olewnik, L. Lung segments from anatomy to surgery. *Folia Morphol. (Warsz.)* **83**, 20–34 (2024).
- Chen, X. et al. A fully automated noncontrast CT 3-D reconstruction algorithm enabled accurate anatomical demonstration for lung segmentectomy. *Thorac. Cancer* **13**, 795–803 (2022).
- Li, X. et al. Accuracy and efficiency of an artificial intelligence-based pulmonary broncho-vascular three-dimensional reconstruction system supporting thoracic surgery: retrospective and prospective validation study. *eBioMedicine* **87**, 104422 (2023).
- McDonald, M. & D. Shirk, J. Application of three-dimensional virtual reality models to improve the pre-surgical plan for robotic partial nephrectomy. *JSLs: J. Soc. Laparosc. Robotic Surg.* **25**, e2021.00011 (2021).
- Robb, H. et al. The current and possible future role of 3D modelling within oesophagogastric surgery: a scoping review. *Surg. Endosc.* **36**, 5907–5920 (2022).
- Sighinolfi, M. C. et al. Three-dimensional customized imaging reconstruction for urological surgery: diffusion and role in real-life practice from an international survey. *J. Pers. Med.* **13**, 1435 (2023).
- Song, C. et al. Use of individualized 3D-printed models of pancreatic cancer to improve surgeons' anatomic understanding and surgical planning. *Eur. Radio.* **33**, 7646–7655 (2023).
- Chen, X. et al. AI-based chest CT semantic segmentation algorithm enables semi-automated lung cancer surgery planning by recognizing anatomical variants of pulmonary vessels. *Front. Oncol.* **12**, 1021084 (2022).
- Wang, X. et al. Application of three-dimensional (3D) reconstruction in the treatment of video-assisted thoracoscopic complex segmentectomy of the lower lung lobe: a retrospective study. *Front Surg.* **9**, 968199 (2022).
- Chen, K. et al. Three-dimensional reconstruction computed tomography in thoracoscopic segmentectomy: a randomized controlled trial. *Eur. J. Cardiothorac. Surg.* **66**, ezae250 (2024).
- Chen, X. et al. A fully automated noncontrast CT 3-D reconstruction algorithm enabled accurate anatomical demonstration for lung segmentectomy. *Thorac. Cancer* **13**, 795–803 (2022).
- Matsumoto, T. et al. Comparison of three software programs for three-dimensional graphic imaging as contrasted with operative findings. *Eur. J. Cardiothorac. Surg.* **41**, 1098–1103 (2012).
- Qiu, D., Cheng, Y. & Wang, X. Medical image super-resolution reconstruction algorithms based on deep learning: a survey. *Comput Methods Prog. Biomed.* **238**, 107590 (2023).
- Topal, E., Löffler, M. & Zschech, E. Deep learning-based inaccuracy compensation in reconstruction of high resolution XCT data. *Sci. Rep.* **10**, 7682 (2020).
- Wang, J., Liang, J., Cheng, J., Guo, Y. & Zeng, L. Deep learning based image reconstruction algorithm for limited-angle translational computed tomography. *PLoS One* **15**, e0226963 (2020).
- Kiani, A. et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *npj Digital Med.* **3**, 23 (2020).
- Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat. Med.* **28**, 31–38 (2022).
- Tschandl, P. et al. Human–computer collaboration for skin cancer recognition. *Nat. Med.* **26**, 1229–1234 (2020).
- Liu, K. et al. Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance. *Radiology: Artif. Intell.* **1**, e180084 (2019).
- Mei, J., Cheng, M. M., Xu, G., Wan, L. R. & Zhang, H. SANet: a slice-aware network for pulmonary nodule detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 4374–4387 (2022).
- Park, A. et al. Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet model. *JAMA Netw. Open* **2**, e195600 (2019).
- Wang, Y. et al. IILS: Intelligent imaging layout system for automatic imaging report standardization and intra-interdisciplinary clinical workflow optimization. *EBioMedicine* **44**, 162–181 (2019).
- Patel, B. N. et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis. *NPJ Digit. Med.* **2**, 111 (2019).
- Obuchowski, N. A. & Bullen, J. Multireader diagnostic accuracy imaging studies: fundamentals of design and analysis. *Radiology* **303**, 26–34 (2022).
- Schalekamp, S., van Ginneken, B., Schaefer-Prokop, C. M. & Karssemeijer, N. Influence of study design in receiver operating characteristics studies: sequential versus independent reading. *J. Med. Imaging (Bellingham)* **1**, 015501 (2014).

Acknowledgements

This study is funded by the Ministry of Science and Technology of the People's Republic of China. Funding No. 2020AAA0109600 (Jun Wang)

and 2021ZD011104 (Dawei Wang); CAMS Medical and Health Science and Technology Innovation Project (2021-I2M-5-002) (Jun Wang); National Natural Science Foundation of China (No. 92059203 (Jun Wang), 82388102 (Jun Wang)), Research Unit of Intelligence Diagnosis and Treatment in Early Non-small Cell Lung Cancer, Chinese Academy of Medical Sciences (2021RU002) (Jun Wang), National Natural Science Foundation of China Youth Fund (82002983) (Xiuyuan Chen), grant from Shanghai Hospital Development Center (SHDC1202111) (Yuming Zhu), Hunan Provincial Key Area R&D Program (2023SK2034) (Fenglei Yu). This study is also funded by Intervision Medical Technology Co., Ltd (Fan Yang, Yuming Zhu, Fenglei Yu). We thank Prof. Yanxia Luo from the Statistics Teaching and Research Department of Capital Medical University for providing data entry and management of the MRMC study. We thank Prof. Xiaohua Zhou from the Biostatistics Department of the School of Public Health of Peking University for giving independent data analysis reports. We thank Dr. Yazhen Yin and Dr. Jiarui Sun for their input in the design and analysis methodology of the MRMC study.

Author contributions

X.C. study design, drafting manuscript, results analyzing, MRMC reader; C.D. drafting manuscript, result analyzing, data collection; M.P. drafting manuscript, result analyzing, MRMC reader; D.W. study design, drafting manuscript, results analyzing; X.S. gold standard; L.D. gold standard; X.W. gold standard; X.W. MRMC reader; W.W. MRMC reader; S.W. MRMC reader; H.Z. data cleaning; Z.W. data cleaning; J.G. result analyzing; C.C. MRMC reader; Y.H. MRMC reader; Q.H. data collection; C.J. MRMC reader; H.Z. MRMC reader; Y.B. MRMC reader; C.S. data collection; Z.C. result analyzing; X.Z. result analyzing; H.H. result analyzing; C.X. draft revision, study design; J.L. MRMC coordinator; B.Y. MRMC coordinator; Ji.Qi. MRMC coordinator; F.Ji. preparation for testing virtual machine; S.W. draft revision; K.C. draft revision, study design and technical support as senior author; N.H. draft revision; J.W. draft revision; Y.Z. study design, draft revision; F.Yu. study design, draft revision; F.Y. study design, draft revision.

Competing interests

Dawei Wang, Chen Xia, Jinlong Liu, Bing Yang, Ji Qi, Fanghang Ji, Shaokang Wang are paid employees of Intervision Medical Technology Co., Ltd. The research was partially funded by Intervision Medical

Technology Co., Ltd. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-59200-8>.

Correspondence and requests for materials should be addressed to Yuming Zhu, Fenglei Yu or Fan Yang.

Peer review information *Nature Communications* thanks Clemens Aigner, Alessio Crippa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Thoracic Surgery, Peking University People's Hospital, Beijing, China. ²Thoracic Oncology Institute, Peking University People's Hospital, Beijing, China. ³Research Unit of Intelligence Diagnosis and Treatment in Early Non-small Cell Lung Cancer, Chinese Academy of Medical Sciences, Peking University People's Hospital, Beijing, China. ⁴Institute of Advanced Clinical Medicine, Peking University, Beijing, China. ⁵Beijing Key Laboratory of Innovative Application of Big Data in Lung Cancer, Peking University People's Hospital, Beijing, China. ⁶Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China. ⁷Department of Thoracic Surgery, The Second Xiangya Hospital of Central South University, Changsha, Hunan, China. ⁸Institute of Advanced Research, Intervision Medical Technology Co., Ltd, Beijing, China. ⁹Department of Radiology, Peking University People's Hospital, Beijing, China. ¹⁰These authors contributed equally: Xiuyuan Chen, Chenyang Dai, Muyun Peng, Dawei Wang. ¹¹These authors jointly supervised this work: Yuming Zhu, Fenglei Yu, Fan Yang. ✉ e-mail: drzhuyuming@163.com; yufenglei@csu.edu.cn; dr.yangf@gmail.com