Artificial Intelligence Assisted Real Time Recognition of Intra Abdominal Metastasis during Laparoscopic Gastric Cancer Surgery

Supplementary Material

Table of Content

Supplementary Note 1

In the laparoscopic exploration (LE) videos for advanced gastric cancer (GC), the intra-abdominal metastasis (IAM) lesions could be categorized in different aspects:

(1). Metastatic location: peritoneum, omentum, bowels, mesentery, liver surface and uterus.

(2). Metastatic extent: single, multiple and extensive. Based on our surgical practice, we define a frame with only one lesion as “single”. When there are several lesions present in the frame that are not confluence, we define the frame as “multiple”. If the frame contains large and confluence lesions, or the lesions are present on multiple structures and organs, we define the frame as “extensive”.

(3). Lesion size: tiny and non-tiny lesions. Currently, the Peritoneal Cancer Index (PCI) is recommended for evaluating the peritoneal metastatic tumor burden according to the Chicago Consensus on Peritoneal Surface Malignancies[1]. In PCI assessment, the IAM lesion size (LS) score is categorized into four levels: (1) LS score = 0: no tumor; (2) LS score = 1: size ≤ 0.5 cm; (3) LS score = 2: 0.5 cm < size ≤ 5.0 cm; (4) LS score = 3: size > 5.0cm or confluence. Based on existing literature and clinical practice in diagnostic laparoscopy, we consider lesions with diameter ≤ 0.5cm (LS score = 1) as “tiny lesions” that are indeed prone to being overlooked. There is no appropriate measurement tool in laparoscopy. And lesions’ shape would be change after resection, which can lead to inaccurate size measurements. As a result, lesion diameter is typically assessed through observation and estimated using the tip of instruments, such as laparoscopic gripper (about 0.5 cm) during diagnostic laparoscopy[2] (Supplementary Figure 4). Therefore, we define lesions meeting these criteria as “tiny lesions”: (1) lesions with diameter ≤ 0.5cm, meanwhile, similar to or smaller than the tip of the surgical instruments; (2) presence of only a single lesion in the scene.

Supplementary Note 2

The annotation protocol is detailed as follows:

(1). Annotate only intact and clearly visible IAM lesions.

(2). Avoid annotating：

1） Areas substantially covered with smoke/blood/fat;

2） Areas not properly visible due to soiling of the laparoscope;

3） Devices such as surgical instrument, gauze, needle, etc.;

4） Any areas outside the image margins.

(3). Annotate IAM lesions whenever it is possible to recognize them in an area that is：

1） Dark or reflective;

2） Slightly covered with smoke;

3） Small: sometimes, parts of the lesions are visible in tiny areas, such as in instrument;

(4). Note that IAM lesions may be visible in several small areas in a single image.

(5). Note that IAM lesions may be visible in different abdominal regions including peritoneum, omentum, bowels, mesentery, liver surface, uterus, adnexa, etc.

(6). Note that IAM lesions have different extents including single, multiple and extensive, which have differences in annotation process (Supplementary Figure 5).

Supplementary Note 3

As a supplement to the setting of deep learning models in the manuscript. The details are provided here for reference:

(1). Artificial intelligence laparoscopic exploration system (AiLES)

The architecture of AiLES was based on the Residual feedback network (RF-net)[3]. The network includes two steps: (1) in the first step, an encoder-decoder architecture is used to create initial segmentation outcomes from the input lesion images. A residual representation module is then applied, which processes the decoder block features to capture information about low-confidence areas and incorrect pixel predictions. This step is regulated by residual masks, which highlight the discrepancies between the ground truth and the initial segmentation. (2) In the second step, the representation module is used to correct errors through residual feedback transmission strategy. The encoder-decoder framework is then reused to refine and generate improved segmentation results based on this residual guidance.

Weighted-balanced and weighted binary cross-entropy were used as loss function. The loss was updated for a maximum of 150 epochs using the stochastic gradient descent (SGD) optimizer (momentum=0.9, decay=0.01) with a base learning rate of 0.001. If the validation loss did not show any improvement over a span of 10 consecutive epochs, the learning rate was reduced by half, and the model was set to stop training early if no improvement occurred for another 10 consecutive epochs. The model parameters from the epoch with the last observed improvement in validation loss were saved.

(2). DeeplabV3+ model

We adopted Xception as the backbone for the DeeplabV3+ deep learning model[4]. Binary cross-entropy were used as loss function. The loss was updated for a maximum of 150 epochs using the SGD optimizer (momentum=0.9, decay=0.01) with a base learning rate of 0.001. If the validation loss did not show any improvement over a span of 10 consecutive epochs, the learning rate was reduced by half, and the model was set to stop training early if no improvement occurred for another 10 consecutive epochs. The model parameters from the epoch with the last observed improvement in validation loss were saved.

(3). Segment Anything Model (SAM)

Segment Anything Model is the first universal image segmentation foundation model that aims at segmenting objects using prompts[5]. These prompts could be a single point, multiple points (including full masks), bounding boxes, or text descriptions. We used the pre-trained "ViT-Base" model as the image encoder. For our test dataset, we evaluated the performance of SAM by creating one-point prompt and one-box prompt per image and then evaluating the predicated segmentation accuracy by comparing to the "ground truth" mask annotations. Furthermore, we conducted tests to evaluate the performance of the SAM in automated segmentation.

(4). Medical SAM Adapter (MSA)

The Medical SAM Adapter (MSA) demonstrated outstanding performance across 19 medical image segmentation tasks involving various imaging modalities such as computed tomography, magnetic resonance imaging, ultrasound, fundus, and dermoscopic images, surpassing the performance of the original SAM[6]. This improvement was achieved by pre-training the model encoder specifically with medical images. We used the pre-trained "ViT-Base" model as the image encoder and fine-turned the MSA with prompt on our training set. As same as the SAM, we conducted tests to evaluate the segmentation performance of the MSA in automatic segmentation mode.

Supplementary Note 4

The detailed metrics for model performance evaluation:

(1). For segmentation task, Dice score is a measure of overlap between prediction and ground truth[7], while intersection-over-union (IOU) evaluates the accuracy of a segmentation by comparing the area of overlap to the area of union. The formulas of Dice and IOU are detailed below:

$$\text{Dice}=\frac{2\times|A\cap B|}{|A|+|B|}$$

$$\text{IOU (Jaccard index)}=\frac{|A\cap B|}{|A\cup B|}$$

Note: A denotes the segmentation predicted by the algorithms, while B refers to the manually annotated reference segmentation.

(2). The metrics including accuracy, precision, recall (sensitivity), specificity and F1 score are pixel-level evaluation metrics in the segmentation tasks. The formulas are detailed below:

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}}$$

$$\text{Recall (Sensitivity)}=\frac{\text{TP}}{\text{TP+FN}}$$

$$\text{Specificity}=\frac{\text{TN}}{\text{TN+FP}}$$

$$\text{F1 score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}}=\frac{2\times\text{TP}}{2\times\text{TP+FP+FN}}=\text{Dice score}$$

Note: TP (true positive); FP (false positive); TN (true negative); FN (false negative). Additionally, it should be noted that F1 score and Dice score are interchangeable in image segmentation tasks, yielding the same numerical results.

(3). Mean average precision at IOU of 50% (mAP@50) quantifies the mean average precision when the IOU between predicted results and ground truth annotations reaches 50%[8]. Specifically, mAP@50 computes the Average Precision (AP) for each class, which is the mean precision at various recall levels, and then averages these AP values across all classes to yield mAP@50. The mAP is computed by averaging the AP values across all classes in the dataset. In this study, as the only object to be segmented is the IAM lesion, mAP@50 is equivalent to AP@50 of lesions.

(4). In this study, similarity indices include Structural similarity index measure (SSIM), Hausdorff distance (HD), Dice and IOU. SSIM could assess the structural similarity between segmentation mask and ground truth (GT). HD could evaluate the similarity of point sets from segmentation mask and ground truth. Also, Dice and IOU are set similarity metrics. The formulas are displayed below:

$$\mathrm{HD}\ (A,B)=\max(\max_{a\in A}\{\min_{b\in B} d(a,b)\}, \max_{b\in B}\{\min_{a\in A} d(a,b)\})$$

Note: A denotes the segmentation predicted by the algorithms, while B refers to the manually annotated reference segmentation. A={a1,a2,...,am} and B={b1,b2,...,bn}, d(a,b)d(a,b) represents the distance between points a and b.

$$\mathrm{SSIM}\ (A,B)=\frac{(2\mu_A\mu_B+c_1)(\sigma_{AB}+c_2)}{(\mu_A^2+\mu_B^2+c_1)(\sigma_A^2+\sigma_B^2+c_2)}$$

Note: A denotes the segmentation predicted by the algorithms, while B refers to the manually annotated reference segmentation. μ_A and μ_B are the mean intensities of images A and B, respectively. σ_A^2 and σ_B^2 are the variances of images A and B, respectively. σ_{AB} is the covariance of images A and B. c_1 and c_2 are constants used *to stabilize the division with weak denominators. The SSIM value ranges between 0 and 1, where values closer to 1 denote a greater similarity between the two images.*

Supplementary References

1. The Chicago Consensus on Peritoneal Surface Malignancies: Management of Gastric Metastases. *Ann Surg Oncol* **27**, 1768-1773 (2020).

2. Bresson L*, et al.* Single-port or Classic Laparoscopy Compared With Laparotomy to Assess the Peritoneal Cancer Index in Primary Advanced Epithelial Ovarian Cancer. *J Minim Invasive Gynecol* **23**, 825-832 (2016).

3. Wang K, Liang S, Zhang Y. Residual Feedback Network for Breast Lesion Segmentation in Ultrasound Image. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*) (2021).

4. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: *European Conference on Computer Vision*) (2018).

5. Alexander Kirillov*, et al.* Segment Anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015-4026 (2023).

6. Wu J*, et al.* Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation. *ArXiv* **abs/2304.12620**, (2023).

7. Kolbinger FR*, et al.* Anatomy segmentation in laparoscopic surgery: comparison of machine learning and human expertise - an experimental study. *Int J Surg* **109**, 2962-2974 (2023).

8. Bolya D, Zhou C, Xiao F, Lee YJ. YOLACT: Real-Time Instance Segmentation. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*) (2019).

Supplementary Table 1. Patient characteristics and lesion distribution according to metastatic extents and locations in the whole dataset.

Characteristics	Value
Number of Patients	100
Age(years) *	62.4±6.5
BMI (kg/m^2) *	24.2±4.1
Sex	
Female	54(54.0%)
Male	46(46.0%)
Number of Images	5111
Metastatic extent	
Single	2254(44.1%)
Multiple	1798(35.2%)
Extensive	1059(20.7%)
Metastatic location	
Peritoneum	3221(63.0%)
Omentum	668(13.1%)
Bowels	186(3.6%)
Mesentery	124(2.4%)
Liver surface	819(16.0%)
Uterus	93(1.9%)

*For age and BMI, data were expressed in mean (±standard deviation, SD). BMI: Body Mass Index.

Supplementary Table 2. The similarity index table.

Model	**Dice score↑**	**IOU↑**	**SSIM↑**	**HD↓**
SAM-Anything	0.14(0.30)	0.07(0.28)	0.92(0.11)	264.59(312.58)
SAM-box	0.29(0.32)	0.17(0.31)	0.96(0.07)	172.23(177.81)
SAM-point	0.02(0.10)	0.01(0.07)	0.80(0.23)	1080.83(343.08)
MSA	0.63(0.31)	0.46(0.29)	0.99(0.02)	105.43(106.14)
DeeplabV3+	0.67(0.14)	0.50(0.13)	0.99(0.01)	95.62(81.95)
AiLES	**0.76(0.17)**	**0.61(0.19)**	**0.99(0.01)**	**67.88(40.82)**

Data were expressed in mean (±standard deviation, SD). Higher Dice score, IOU, SSIM and lower HD indicate better performance. IOU: intersection-over-union (also called Jaccard index); SSIM: structural similarity index measure; HD: Hausdorff distance; SAM: Segment Anything Model; MSA: Medical SAM Adapter; AiLES: artificial intelligence laparoscopic exploration system.

Supplementary Table 3. The inference speed of different models.

Model	**Inference speed (GPU×2)**		**Inference speed (GPU×1)**	
	Model inference	**Whole process**	**Model inference**	**Whole process**
SAM	4 fps	4 fps	2 fps	2 fps
MSA	4 fps	4 fps	2 fps	2 fps
DeeplabV3+	13 fps	10 fps	8 fps	7 fps
AiLES	**27 fps**	**17 fps**	**15 fps**	**11 fps**

The whole process includes image loading, model inference and prediction results visualization. SAM: Segment Anything Model; MSA: Medical SAM Adapter; AiLES: artificial intelligence laparoscopic exploration system; fps: frames per second.

Supplementary Table 4. Number of videos and frames used in surgical artificial intelligence segmentation studies.

Study title	**Surgical Procedure**	**Video**	**Frame**	**Segmentation object**
Deep-learning-based semantic segmentation of autonomic nerves from laparoscopic images of colorectal surgery: an experimental pilot study	Laparoscopic left-sided colorectal resections	245	12978	Nerves
Artificial intelligence for the recognition of key anatomical structures in laparoscopic colorectal surge	laparoscopic colorectal resections	252	10711	Ureter and nerves
Our study	**Laparoscopic exploration for gastric cancer**	**100**	**5111**	**Intra-abdominal metastasis**
Vessel and tissue recognition during third-space endoscopy using a deep learning algorithm	Endoscopic submucosal dissection	16	2012	Vessel, tissue and instrument
Precise highlighting of the pancreas by semantic segmentation during robot-assisted gastrectomy: visual assistance with artificial intelligence for surgeons	Robot-assisted gastrectomy	62	1158	Pancreas
Deep learning-based recognition of key anatomical structures during robot-assisted minimally invasive esophagectomy	Robot-assisted minimally invasive esophagectomy	83	1050	Azygos vein, vena cava, aorta and lung

Supplementary Table 5. TRIPOD+AI Checklist. The checklist of Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (Artificial Intelligence).

Section/Topic	Item	Development / evaluation[1]	Checklist item	Reported on page
TITLE				
Title	1	D;E	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	1
ABSTRACT				
Abstract	2	D;E	See TRIPOD+AI for Abstracts checklist	3
INTRODUCTION				
Background	3a	D;E	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	4
	3b	D;E	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (e.g., healthcare professionals, patients, public)	4-5
	3c	D;E	Describe any known health inequalities between sociodemographic groups	-
Objectives	4	D;E	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	6
METHODS				
Data	5a	D;E	Describe the sources of data separately for the development and evaluation datasets (e.g., randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	19-20
	5b	D;E	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	20
Participants	6a	D;E	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including the number and location of centres	20
	6b	D;E	Describe the eligibility criteria for study participants	19-20
	6c	D;E	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	19, Table 1
Data preparation	7	D;E	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	20-23
Outcome	8a	D;E	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	19-22
	8b	D;E	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	-
	8c	D;E	Report any actions to blind assessment of the outcome to be predicted	-
Predictors	9a	D	Describe the choice of initial predictors (e.g., literature, previous models, all available predictors) and any pre-selection of predictors before model building	20
	9b	D;E	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	20
	9c	D;E	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	-
Sample size	10	D;E	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	20
Missing data	11	D;E	Describe how missing data were handled. Provide reasons for omitting any data	7, 20
Analytical methods	12a	D	Describe how the data were used (e.g., for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	20-22
	12b	D	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation).	22
	12c	D	Specify the type of model, rationale[2], all model-building steps, including any hyperparameter tuning, and method for internal validation	23-24, Supplementary Note 3
	12d	D;E	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (e.g., hospitals, countries). See TRIPOD-Cluster for additional considerations[3]	-
	12e	D;E	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	24-25, Figure 1
	12f	E	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	-

	12g	E	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	24-25, Supplementary Note 4
Class imbalance	13	D;E	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	7
Fairness	14	D;E	Describe any approaches that were used to address model fairness and their rationale	-
Model output	15	D	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	24-25
Training versus evaluation	16	D;E	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	24
Ethical approval	17	D;E	Name the institutional research board or ethics committee that approved the study and describe the participant-informed consent or the ethics committee waiver of informed consent	19
OPEN SCIENCE				
Funding	18a	D;E	Give the source of funding and the role of the funders for the present study	26
Conflicts of interest	18b	D;E	Declare any conflicts of interest and financial disclosures for all authors	27
Protocol	18c	D;E	Indicate where the study protocol can be accessed or state that a protocol was not prepared	-
Registration	18d	D;E	Provide registration information for the study, including register name and registration number, or state that the study was not registered	19
Data sharing	18e	D;E	Provide details of the availability of the study data	25
Code sharing	18f	D;E	Provide details of the availability of the analytical code[4]	25-26
PATIENT & PUBLIC INVOLVEMENT				
Patient & Public Involvement	19	D;E	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement.	-
RESULTS				
Participants	20a	D;E	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	6, Figure 1
	20b	D;E	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups.	6, Figure 2, Supplementary Table 1
	20c	E	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome).	6, Supplementary Table 1
Model development	21	D;E	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	6-7
Model specification	22	D	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third-party evaluation and implementation, including any restrictions to access or re-use (e.g., freely available, proprietary)[5]	24-25, Supplementary Note 4
Model performance	23a	D;E	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation.	8-9, Table 3, Table 4, Figure 3, Figure 4, Figure 5, Supplementary Table 2, Supplementary Table 3
	23b	D;E	If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD Cluster for additional details[3].	-
Model updating	24	E	Report the results from any model updating, including the updated model and subsequent performance	-
DISCUSSION				
Interpretation	25	D;E	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	10-17

Limitations	26	D;E	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalizability	17-18
Usability of the model in the context of current care	27a	D	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	7, 12-13
	27b	D	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	16-17
	27c	D;E	Discuss any next steps for future research, with a specific view to applicability and generalizability of the model	18

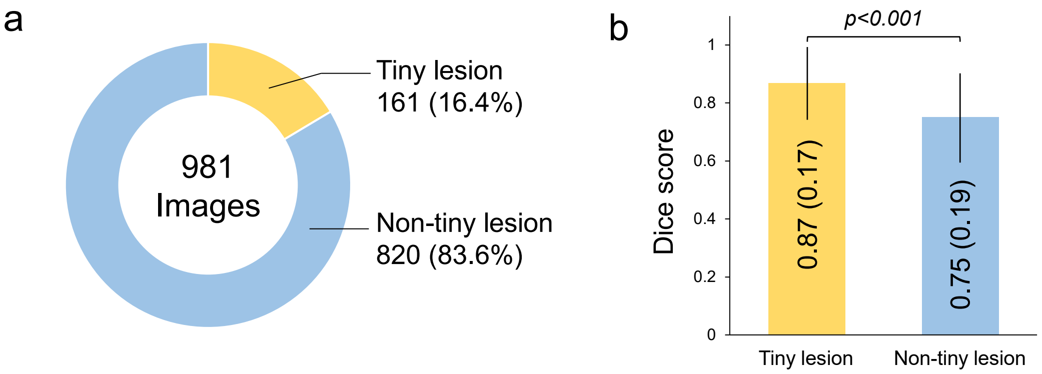
[1] D=items relevant only to the development of a prediction model; E=items relating solely to the evaluation of a prediction model; D;E=items applicable to both the development and evaluation of a prediction model

[2] Separately for all model building approaches.

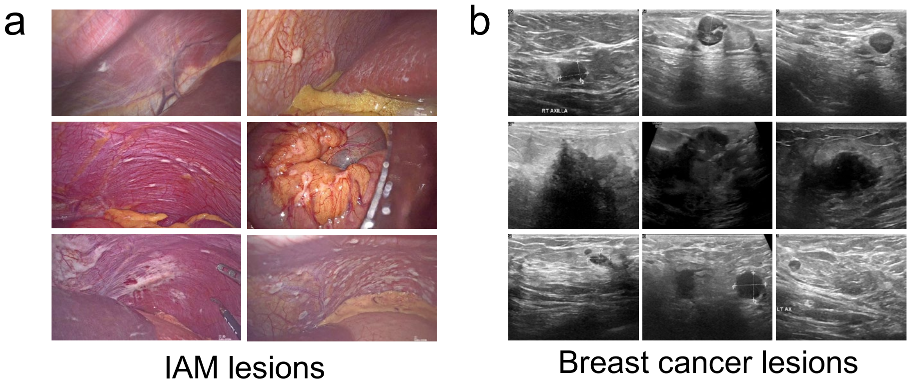
[3] TRIPOD-Cluster is a checklist of reporting recommendations for studies developing or validating models that explicitly account for clustering or explore heterogeneity in model performance (eg, at different hospitals or centres). Debray et al, BMJ 2023; 380: e071018 [DOI: 10.1136/bmj-2022-071018]

[4] This relates to the analysis code, for example, any data cleaning, feature engineering, model building, evaluation.

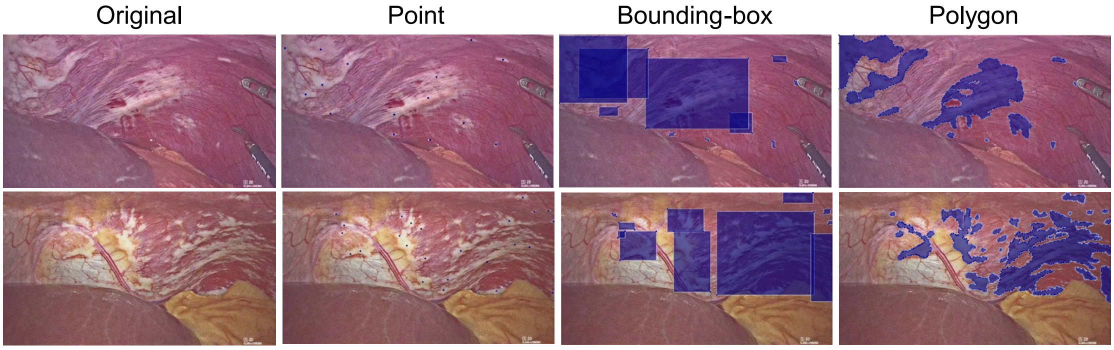
[5] This relates to the code to implement the model to get estimates of risk for a new individual.



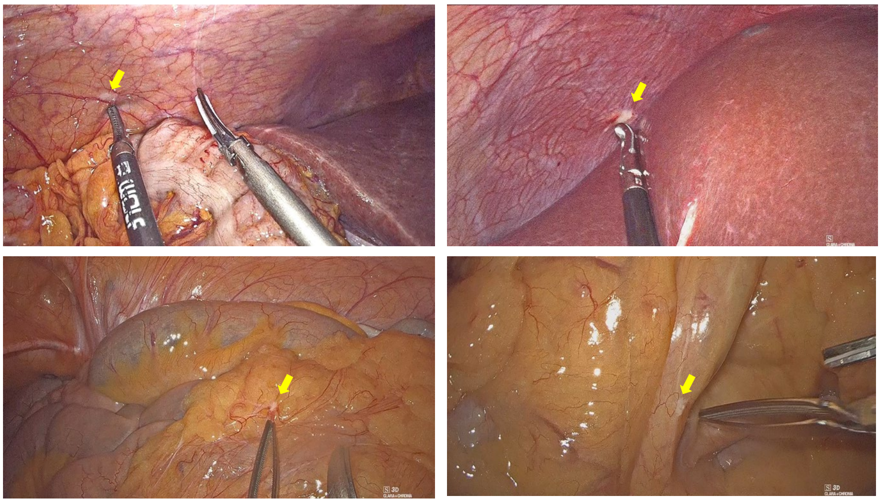
Supplementary Figure 1. Performance evaluation of AiLES on tiny lesions. a. Data percentage of tiny lesions in test dataset; **b.** Segmentation performance on tiny and non-tiny lesions by AiLES. AiLES: artificial intelligence laparoscopic exploration system.



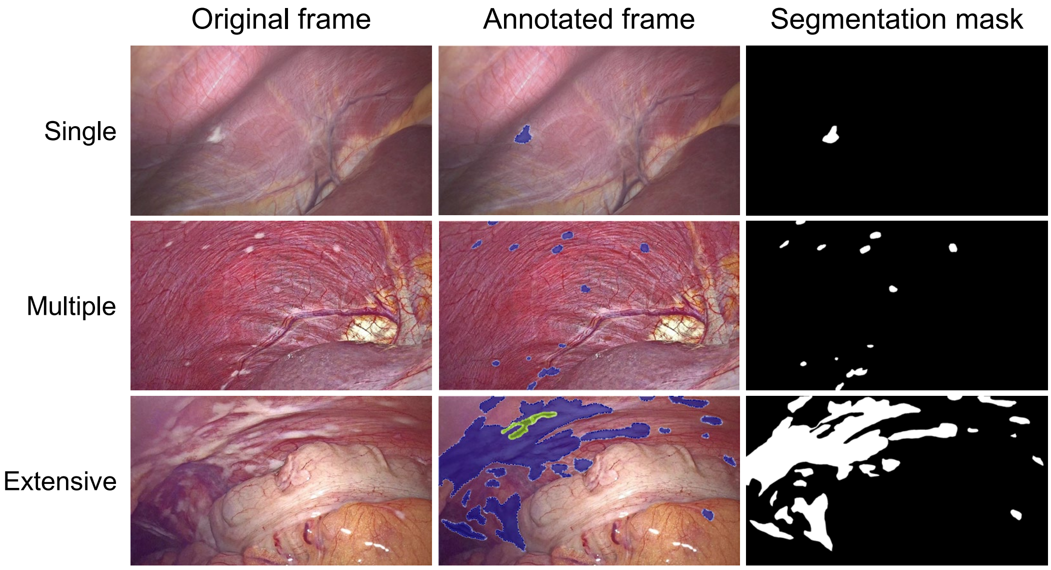
Supplementary Figure 2. Different types of medical image data. a. Laparoscopic images of intra-abdominal metastasis lesions (Nanfang Hospital, China); **b.** Ultrasound images of breast cancer lesions (Baheya Hospital, Egypt).



Supplementary Figure 3. Different annotation approaches (point, bounding-box and polygon) and visual effect of IAM annotation. The comparison of various annotation approaches demonstrated that the polygon approach is most suitable for annotating intra-abdominal metastasis lesions, as it accurately outlines lesion boundaries regardless of their shape or extent. IAM: intra-abdominal metastasis.



Supplementary Figure 4. Using tips of instruments as the tool to estimate tiny lesion diameter. In this study, lesions with a diameter similar or lower to 0.5 cm were defined as tiny lesions. There is no appropriate measurement tool in laparoscopy. And lesions' shape would change after resection, which would lead to inaccurate size measurements. As a result, lesion diameter is typically assessed through observation and estimated using the tip of instruments, such as laparoscopic gripper (about 0.5 cm) during laparoscopic exploration.



Supplementary Figure 5. Annotation samples of different cases of single, multiple and extensive metastasis. In the column of annotated frame, the blue annotation refers to the metastasis, the green annotation refers to the normal structures or tissues surrounded by lesions.

Supplementary Movie 1. Real-time recognition of intra-abdominal metastasis. This movie presents two cases demonstrating the real-time recognition capabilities of AiLES. **Case 1:** Real-time recognition of single and tiny lesion. **Case 2:** Real-time recognition of lesions with different extents, shapes and boundaries. AiLES: artificial intelligence laparoscopic exploration system.

Supplementary Data 1

Data file 1. Lengths of all videos in the study dataset (Source data of Figure 2a). Original videos include clips of all laparoscopic exploration steps (trocar insertion, intra-abdominal exploration, peritoneal cytology, resection of suspicious lesions with biopsy, and closure of abdominal incisions and others). Edited videos focus only on clips of intra-abdominal exploration.

Data file 2. Number of frames in different categories including metastatic extents and locations (Source data of Figure 2b). Metastatic extents include single, multiple, and extensive. Metastatic locations include peritoneum, omentum, bowels, mesentery, liver surface and uterus.

Data file 3. The performance metrics of novice surgeons and AiLES (Source data of Figure 5a). The metrics include Dice score (same as F1 score), intersection-over-union (IOU), sensitivity (same as recall), specificity, accuracy and precision. AiLES: artificial intelligence laparoscopic exploration system.

Data file 4. The Dice score of novice surgeons and AiLES in recognition of IAM with different metastatic extents and locations (Source data of Figure 5b). Metastatic extents include single, multiple, and extensive. Metastatic locations include peritoneum, omentum, bowels, mesentery, liver surface and uterus. AiLES: artificial intelligence laparoscopic exploration system.

Data file 5. The performance of AiLES in recognition of tiny lesions (Source data of Supplementary Figure 1). This data file includes the number of frames with tiny lesions in test dataset and the Dice score of AiLES in recognition of tiny lesions. AiLES: artificial intelligence laparoscopic exploration system.