

80665\_Auto\_Edited.docx

**Name of Journal:** *World Journal of Gastroenterology*

**Manuscript NO:** 80665

**Manuscript Type:** ORIGINAL ARTICLE

*Retrospective Study*

**MRI-based deep learning model to predict multiple firings in double-stapled colorectal anastomosis**

Deep learning to avoid multiple firings

## **Abstract**

### **BACKGROUND**

Multiple linear stapler firings during double stapling technique (DST) after laparoscopic low anterior resection (LAR) are associated with an increased risk of anastomotic leakage (AL). However, it is difficult to predict preoperatively the need for multiple linear stapler cartridges during DST anastomosis.

### **AIM**

To develop a deep learning model to predict multiple firings during DST anastomosis based on pelvic magnetic resonance imaging (MRI).

### **METHODS**

We collected 9476 MR images from 328 mid-low rectal cancer patients undergoing LAR with DST anastomosis, which were randomly divided into a training set ( $n = 260$ ) and testing set ( $n = 68$ ). Binary logistic regression was adopted to create a clinical model using six factors. The sequence of fast spin-echo (FSE) T2-weighted MRI of the entire pelvis was segmented and analyzed. Pure-image and clinical-image integrated deep learning models were constructed using the mask region-based convolutional neural network segmentation tool and three-dimensional convolutional networks. Sensitivity, specificity, accuracy, positive predictive value (PPV), and area under the receiver operating characteristic curve (AUC) was calculated for each model.

### **RESULTS**

The prevalence of  $\geq 3$  Linear stapler cartridges was 17.7% (58/328). The prevalence of AL was statistically significantly higher in patients with  $\geq 3$  cartridges compared to those with  $\leq 2$  cartridges (25.0% *vs.* 11.8%,  $p = 0.018$ ). Preoperative carcinoembryonic antigen level  $> 5$  ng/mL (OR = 2.11, 95%CI 1.08- 4.12,  $p = 0.028$ ) and tumor size  $\geq 5$  cm (OR = 3.57, 95%CI 1.61 - 7.89,  $p = 0.002$ ) were recognized as independent risk factors for use of  $\geq 3$  Linear stapler cartridges. Diagnostic performance was better with the

integrated model (accuracy = 94.1%, PPV = 87.5%, and AUC = 0.88) compared with the clinical model (accuracy = 86.7%, PPV = 38.9%, and AUC = 0.72) and the image model (accuracy = 91.2%, PPV = 83.3%, and AUC = 0.81).

## CONCLUSION

MRI-based deep learning model can predict the use of  $\geq 3$  Linear stapler cartridges during DST anastomosis in laparoscopic LAR surgery. This model might help determine the best anastomosis strategy by avoiding DST when there is a high probability of the need for  $\geq 3$  Linear stapler cartridges.

**Key Words:** Deep learning; Image-reading artificial intelligence; Magnetic resonance imaging; Predictive model; Double stapling technique; Linear stapler; Rectal cancer; Laparoscopic surgery; Low anterior resection; Anastomotic leakage

Cai ZH, Zhang Q, Fu ZW, Fingerhut A, Tan JW, Zang L, Dong F, Li SC, Wang SL, Ma JJ. MRI-based deep learning model to predict multiple firings in double-stapled colorectal anastomosis. *World J Gastroenterol* 2022; In press

**Core Tip:** Multiple linear stapler firings during double stapling technique (DST) anastomosis is associated with an increased risk of anastomotic leakage after laparoscopic low anterior resection. This retrospective study developed a deep learning model to predict the use of  $\geq 3$  Linear stapler cartridges during DST anastomosis. With the help of the artificial intelligence to identify and extract information from pelvic magnetic resonance imaging, we developed a clinical-image integrated model with satisfactory accuracy. This model might help preoperatively to determine the anastomosis strategy for rectal cancer patients (suggesting not to perform DST when the risk for  $\geq 3$  firings is high).

## INTRODUCTION

Anastomotic leakage (AL) is the most common postoperative complication after laparoscopic low anterior resection (LAR) for mid and low rectal cancer [1]. The consequences of AL include higher mortality, need for remedial re-operation, unplanned stoma, delay before adjuvant therapy, and compromised long-term oncological outcomes [2-4]. Although several techniques have been designed to prevent AL [5-9], the prevalence of this complication has hardly improved over the past 20 years [10,11].

Of these techniques, the double stapling technique (DST) has facilitated bowel reconstruction but failed to eliminate AL [2]. During this procedure, the distal margin of the tumor-bearing specimen is transected by one or more linear stapler firings to create the rectal stump. Several publications have identified multiple linear stapler firings as an independent risk factor for AL [1,6,12-16]. Both the Chinese Expert Consensus Statement on the Diagnostic, Prevention and Treatment of the Anastomotic Leakage for Rectal Cancer (2019) and the U.S. Food and Drug Administration have suggested limiting the number of stapler firings to two in the DST procedure [17,18]. A recent review of DST suggested that alternative anastomotic techniques to avoid multiple firings on the rectal stump might lower the AL rate [11].

If the number of stapler cartridges used during surgery were predictable before operation, we could predetermine whether DST would be the ideal method for reconstruction. Several studies have reported the association between pelvimetry findings and the technical difficulties (including the use of  $\geq 3$  Linear stapler cartridges) in LAR for mid-low rectal cancer [19-21]. However, previous studies only considered the dimension of pelvic bone landmarks in pelvimetry but ignored mesorectum thickness, tumor size, or tumoral infiltration to nearby organs (prostate, seminal vesicle, uterus). Based on our subjective experience, we speculated that the narrow (male) pelvis, thick mesorectum, aggressive tumor infiltration, and low transection margin might be associated with the need for  $\geq 3$  Linear stapler cartridges to close the rectal stump. Besides, a simple comparison of one or several measurements of pelvimetry is insufficient to reveal the difficulty of the pelvic procedure. For a lean female patient or a

heavy male patient, the same interspinous distance has a vast difference in clinical significance. Furthermore, manual measurement of pelvimetry indicators is time-consuming and labor-intensive.

These shortcomings of existing predictive methods prompted us to design and develop a new model to predict more precisely and effectively the need for  $\geq 3$  Linear stapler firings during DST. Pelvic magnetic resonance imaging (MRI), a routine and first-choice tool for preoperative staging of rectal cancer [22], can capture mesorectal or nearby tissue infiltration characteristics in addition to bony structures. On the other hand, machine learning and deep learning models have been widely applied in health care because of their high ability to predict and make decisions [23]. Owing to the recent technological development [24-25], image-reading artificial intelligence (AI) programs can be used to recognize target features, and then interpret images or provide diagnoses based on these target features [26-30].

In this study, we aimed to create a deep learning pre-warning model for the use of multiple linear stapler cartridges during DST anastomosis by adopting AI to identify, extract and integrate image information from pelvic MRI.

## **MATERIALS AND METHODS**

### ***Patients***

We retrospectively analyzed the records of 328 patients who underwent laparoscopic LAR for mid-low rectal cancer at Ruijin Hospital, Shanghai, China, between 2016 and 2021. Clinicopathological data were collected from our prospective institutional database and the study was approved by Ruijin Hospital Ethics Committee (No. 2019-82). Informed consent was waived by the committee because of the retrospective nature of the study. The study was registered at clinicaltrials.gov with the registry number: NCT05498506.

The inclusion criteria were: 1) rectal carcinoma confirmed by histopathological evaluation; 2) tumor located in the mid-low rectum ( $< 10\text{cm}$  from the anal verge); 3) performance of DST anastomosis; 4) pelvic MRI obtained within 14 days before surgery.

The exclusion criteria were: 1) other anastomotic techniques (*e.g.*, trans-anal rectal excision); 2) Hartmann's operation or other procedures without anastomosis; 3) robotic surgery; 4) the number of linear stapler cartridges was not traceable in the operative report. By using an unbiased random sampling method with a split ratio of 4:1, the patients were divided into a training set ( $n = 260$ ) and testing set ( $n = 68$ ).

### ***Surgical procedure***

Laparoscopic LAR was performed by one operating team who treated > 200 cases of rectal cancer per year. The surgical procedure followed the national guidelines for laparoscopic radical resection of colorectal cancer (2018 edition). Distal rectal transection was performed with an endoscopic linear stapler (Endo-GIA™ Ultra Universal Stapler Reload with Tri-staple™ Technology; Covidien LLC, Minneapolis, MN, USA), fired manually through the right lower quadrant 12-mm trocar. The 60-mm purple cartridges containing three different staple heights (3.0, 3.5, and 4.0 mm) were routinely used. However, the 45-mm purple cartridges could be used when the stapler could not be placed perpendicularly to the rectum with the 60-mm cartridges.

### ***Clinical variables and clinical model***

We collected and analyzed baseline characteristics (sex, age, body mass index [BMI]), laboratory analysis (hemoglobin, albumin, carcinoembryonic antigen [CEA]), and tumor features (distance from the anal verge, circumferential resection margin [CRM], tumor size, tumor stage). For the clinical model, we created a multivariate binary logistic regression model based on clinical variables that might be associated with the number of linear stapler cartridges during surgery: three binary variables (sex [male, female], CEA level [normal, elevated (> 5 ng/mL)], and CRM [positive, negative]) and three continuous variables (BMI, distance from the anal verge, and tumor size).

### ***MRI protocol and labeling of target region***

Pelvic MRI was performed by a Philips INGENIA™ MR scanner with a field strength of 3.0 T and the patient in the supine position. The scanning parameters included: repetition time = 3565 ms; echo time = 80 ms; layer thickness = 5 mm; image matrix = 312 × 357, field of view = 250 × 340 × 166 mm.

The sequence of fast spin-echo (FSE) T2-weighted MRI with a large field of view with fat suppression obtained in the axial plane of the entire pelvis was retrieved from the Picture Archiving and Communication System for image segmentation. A total of 9476 T2-weighted MR images were collected from the enrolled patients. Fifteen patients in the training set were randomly selected by random number tables and 367 images from these patients served for manual labeling. A radiological expert with > 15 years of experience in pelvic MRI labeled three target regions (pelvis, mesorectum, and tumor body) on each of the consecutive T2-weighted images. These regions were represented by drab, yellow, and green, respectively (Supplementary Figure 1), using an open annotation tool named Labelme\* [31]. Data were transformed into the COCO dataset format [32].

### *Segmentation model*

Mask region-based convolutional neural network (Mask R-CNN) [24] was used to detect and segment the three target regions (Supplementary Figure 2).

The entire Mask R-CNN network was trained on the training set, and the performance of the testing set was evaluated using the mean Average Precision (mAP). When mAP was > 50, we considered the segmentation model to have performed well [24].

To visualize intuitively the segmentation of the target region, 3D Slicer software \*\* was adopted to reconstruct a three-dimensional visualization model for each patient (Figure 1).

### *Deep learning model*

A three-dimensional convolutional networks (C3D)-based model was used to generate the probability of multiple linear stapler cartridges after segmentation [25]. We used all



the images of one patient as the input whereas the output was the probability of  $\geq 3$  Linear stapler cartridges. When the probability was greater than a preset threshold (set to 0.5 empirically), the sample was judged as positive. We trained the C3D network on the training set for 100 epochs and obtained the final C3D model.

Two deep learning models were used in our study, a pure image model using only T2-weighted MR images segmented by Mask R-CNN and an integrated model using MR images as well as six above-mentioned clinical variables. The flow chart of the design of these pre-warning models is shown in Figure 2. Our source code is publicly available<sup>\*\*\*</sup>. Finally, one clinical model and two deep learning models were evaluated on the testing set. A receiver operating characteristic (ROC) curve was plotted for each model. Sensitivity, specificity, accuracy, positive predictive value (PPV), and area under the curve (AUC) were calculated for each curve.  $AUC > 0.70$  indicated an acceptable model.

### *Statistical analysis*

Statistical Package for the Social Sciences (SPSS 13.0, Chicago, IL, USA) was used for statistical analysis. The statistical methods of this study were reviewed by Shuang Wu from China Novartis Institutes for BioMedical Research Co. Ltd. Numerical variables were examined by non-parametric Wilcoxon rank-sum test. Pearson's Chi-Square or Fisher's exact test was adopted to analyze categorical data. Multivariate analysis was performed by binary logistic regression model. The difference was considered statistically significant if two-sided p values were  $<0.05$ .

\* available at [labelme.csail.mit.edu](http://labelme.csail.mit.edu)

\*\* available at [www.slicer.org](http://www.slicer.org)

\*\*\* <https://github.com/suli609/MRI-DST>.

## **RESULTS**

### *Clinicopathological characteristics of patients*

The entire study population included 328 patients, 227 male and 101 female with a median age of 63 (range 24 - 87) years. The prevalence of use of  $\geq 3$  Linear stapler

cartridges was 17.7% (58/328). The training set ( $n = 260$ ) consisted of 48 cases with  $\geq 3$  cartridges and 212 cases with  $\leq 2$  cartridges. The testing set ( $n = 68$ ) consisted of 10 cases with  $\geq 3$  cartridges and 58 cases with  $\leq 2$  cartridges.

When clinicopathological characteristics were compared between the patients with  $\geq 3$  cartridges and those with  $\leq 2$  cartridges in the training set (Table 1), there was no statistically significant difference between the two groups with respect to sex, age, BMI, diabetes mellitus, preoperative CEA serum level, and the percentage of patients undergoing neoadjuvant chemoradiotherapy. No statistically significant difference was found in the distance from tumor to the anal verge, tumor size, tumor stage, operation time, or insufficient distal resection margin ( $\leq 5$  mm). The incidence of AL was statistically significantly higher in the patients with  $\geq 3$  cartridges compared to those with  $\leq 2$  cartridges ( $p = 0.018$ ).

Univariate and multivariate analysis revealed two independent risk factors for use of  $\geq 3$  Linear stapler cartridges: preoperative CEA level  $> 5$  ng/mL (OR = 2.11, 95%CI 1.08 - 4.12,  $p = 0.028$ ) and tumor size  $\geq 5$  cm (OR = 3.57, 95%CI 1.61 - 7.89,  $p = 0.002$ ) (Table 2). All these clinicopathological features were comparable between the training set and testing set (Table 3).

### *Visualization of target regions*

Of the three-dimensional reconstruction models presented in Figure 1, those in Figure 1A, 1C, and 1E were models from patients with the use of  $\geq 3$  Linear stapler cartridges while those in Figure 1B, 1D, and 1F were models from patients with the use of  $\leq 2$  cartridges. Characteristics potentially relevant to the use of  $\geq 3$  cartridges were narrow pelvis (Figure 1A, drab part), thick mesorectum (Figure 1C, yellow part), and large tumor size with low distal margin (Figure 1E, green part), as can be seen in the models in the left column.

### *Performance of pre-warning models*

The mAP of the segmentation model was 57.2 for the object detection task and 53.7 for the instance segmentation task.

The sensitivity, specificity, and accuracy of the clinical model were 70.0%, 81.0%, and 79.4%, respectively (Youden index = 0.51, PPV = 38.9%). The relevant technical indicators of the image model were as follows: sensitivity = 50.0%, specificity = 98.3%, accuracy = 91.2%, Youden index = 0.48, and PPV = 83.3%. The integrated model showed the best pre-warning performance: sensitivity = 70.0%, specificity = 98.3%, accuracy = 94.1%, Youden index = 0.68, and PPV = 87.5%. Finally, the AUC was 0.72, 0.81, and 0.88 for the clinical model, the image model, and the integrated model, respectively (Figure 3).

## **DISCUSSION**

Our deep learning model can predict the probability of using  $\geq 3$  Linear stapler cartridges in the DST anastomosis during laparoscopic LAR surgery. Compared with the clinical model and the pure image model, the integrated model, which combined both the clinical variables and pelvic MR images, had a better Youden index (0.68) and AUC (0.88). Our results suggest that clinical or imaging information alone is insufficient to predict the use of  $\geq 3$  cartridges during surgery and an MRI-based integrated deep learning model might help determine the best anastomotic strategy for mid-low rectal cancer patients.

The safety, feasibility, and oncological outcomes of laparoscopic LAR surgery for mid-low rectal cancer have been confirmed by a series of high-quality randomized controlled trials [33,34]. During laparoscopic LAR, the DST method is considered to be difficult in some patients because the size and angle of linear staplers are limited in laparoscopy [14,35]. Consequently, multiple stapler firings are often needed. Two mechanisms might give rise to AL: either space is left between two adjacent staple lines, or crossing the staple line with another row of staples or crushing the first staple line with the jaws can dislodge, break or deform the staples [6, 12,13,18].

This has prompted surgeons to modify anastomosis techniques, which have been described as follows: transanal transection of the rectal stump with transanal anastomosis [36,37]; intra-luminal transection of the rectal stump with manual purse-string sutures (e.g., TaTME technique) [37,38]; <sup>3</sup>vertical rectal division using a linear stapler after making an additional skin incision above the pubic symphysis [6]; transverse rectal division using a Contour® stapler during laparoscopic surgery [7]; lateralization of the stump by Nelaton catheter pulling method [8]; side-to-end anastomosis (Baker technique) [9]; trans-anal reinforcement of anastomosis [39]; or removing the “dog ears” / crossing staple lines [40,41].

Thus, if there is a high probability of using  $\geq 3$  cartridges according to preoperative data, one of these other anastomosis methods might be more suitable than the DST method. Foo *et al.* reported a pre-warning model to predict the likelihood of transecting the rectum with  $\geq 3$  <sup>2</sup>stapler cartridges, which included the following parameters: sex, pelvic inlet, interspinous distance, intertuberos distance, and tumor height [21]. Two other studies investigated the technical difficulty in LAR surgery with DST anastomosis but they used other indicators, such as operative time, pelvic operative time, blood loss, conversion rate, complications, or specimen quality [42,43]. The factors associated with technical difficulty were BMI, tumor height, interspinous distance, intertubercle distance, pelvic inlet, and pubic tubercle height. The similarity of these studies with ours is that we combined clinical information with pelvic anatomical factors and the pelvimetry was conducted in pelvic MRI. However, the strengths of our pre-warning model are mainly featured as follows. (1) By using AI-based segmentation of images, the pelvimetry is recognized as a whole instead of isolated measurements. (2) All parameters considered in the above-mentioned clinical models (sex, pelvic measurement, BMI, tumor size/height/stage) were synthesized in our image-reading models. This is why we performed segmentation of three different target regions (bony, fatty, and tumoral) in our study. (3) This AI-based pre-warning model can shorten the prediction time to 100ms. The only data needed are six clinical factors and the sequence of FSE T2-weighted MR images.

Compared with other segmentation algorithms, such as faster R-CNN, the implementation process of Mask R-CNN is simpler, and the segmentation accuracy is higher. The mAP achieved by our model met the needs of most application scenarios [24]. The actual segmentation effect is close to the target regions manually segmented by radiologists (Supplementary Figure 2). The C3D network structure has good versatility, and the overheads of training the model are small, which is suitable for scenarios with limited training samples [25,44].

Our study had several limitations. First, the small sample size in the testing set lowered the statistical power of our analysis. With this sample size, the statistical difference between the three ROC curves might have been underestimated. Second, the lack of cases made it impossible to validate this model in an external set. Further prospective multi-center studies are needed to verify the validity of this model. Third, deep learning was only conducted on FSE T2-weighted sequences with specific scanning parameters. Further studies could focus on other MRI sequences or contrast-enhanced MRI. Fourth, the number of cartridges was not the only factor involved in AL. The intersection of staple lines [45], the precompression before stapler firings [2], and the distance between the linear staple line and the circular end-to-end anastomosis [35] might also have been implicated in addition to the number of firings. However, we could not include these factors in our analysis because of the retrospective nature of our study. Finally, apart from those factors mentioned above, the number of linear stapler cartridges depended on other factors that were difficult to assess, such as the proper lateralization of the intestinal tube [8] and the precise placement of the trocar through which the linear stapler was fired [2,35]. Thus, none of our three models achieved 100% accuracy in the testing set. However, the PPV increased to 87.5% in the integrated model compared with 38.9% in the clinical model, indicating that the trans-abdominal DST method would be unsuitable for positive cases predicted by the integrated model.

## **CONCLUSION**

With the goal of predicting the use of  $\geq 3$  Linear stapler cartridges during DST anastomosis in laparoscopic LAR surgery, our pelvic MRI-based deep learning model might be helpful in the preoperative determination of the best anastomosis strategy for mid-low rectal cancer patients, and, in particular, in avoiding the DST technique when there is a high probability of the need for  $\geq 3$  Linear stapler cartridges. In this setting, another anastomotic technique without staple line crossing should be chosen. Larger studies are needed to validate its clinical value and determine if this strategy can help lower the AL rate.

### **ARTICLE HIGHLIGHTS**

#### ***Research background***

The need for multiple ( $\geq 3$ ) linear stapler firings during double stapling technique (DST) is associated with an increased risk of anastomotic leakage (AL) after laparoscopic low anterior resection (LAR).

#### ***Research motivation***

Current methods using clinical data cannot predict precisely the use of  $\geq 3$  Linear stapler firings before surgery.

#### ***Research objectives***

This study aimed to develop a pelvic magnetic resonance imaging (MRI)-based deep learning model to predict the multiple firings during DST anastomosis.

#### ***Research methods***

Clinical data and 9476 MR images from 328 mid-low rectal cancer patients undergoing LAR with DST anastomosis were retrospectively collected. A pure-image model and a clinical-image integrated model were constructed using image-reading deep learning technologies, respectively.

### ***Research results***

The clinical-image integrated model showed better predictive performance compared with the clinical model and the pure image model with the highest accuracy (94.1%) and area under the curve (0.88).

### ***Research conclusions***

Our deep learning model might help determine the anastomosis strategy for mid-low rectal cancer patients (suggesting not to perform the DST when the risk for  $\geq 3$  Linear stapler firings is high).

### ***Research perspectives***

The clinical value of this clinical-image integrated model will be validated in further prospective studies. The incidence of AL is expected to be decreased with this strategy.

### **ACKNOWLEDGEMENTS**

We express our sincere gratitude to Shuang Wu (Statistical programmer) for her technical assistance.

# 3%

SIMILARITY INDEX

### PRIMARY SOURCES

- 1

assets.researchsquare.com  
Internet

21 words — 1%
- 2

Chi Chung Foo, Hing Tsun Hung, Yuen Chi Ho, Wendy Wai Man Lam, Wai Lun Law. "Predicting the level of difficulty of the double-stapling technique in laparoscopic total mesorectal excision", Surgical Endoscopy, 2019  
Crossref

16 words — < 1%
- 3

Kim, J.S.. "Risk Factors for Anastomotic Leakage after Laparoscopic Intracorporeal Colorectal Anastomosis with a Double Stapling Technique", Journal of the American College of Surgeons, 200912  
Crossref

16 words — < 1%
- 4

Maiko Kobayashi, Hayato Kaida, Akihiko Kawahara, Satoshi Hattori et al. "The Relationship Between GLUT-1 and Vascular Endothelial Growth Factor Expression and 18F-FDG Uptake in Esophageal Squamous Cell Cancer Patients", Clinical Nuclear Medicine, 2012  
Crossref

16 words — < 1%
- 5

www.ncbi.nlm.nih.gov  
Internet

16 words — < 1%
- 6

id.nii.ac.jp  
Internet

14 words — < 1%



7

Hui Qu, Yao Liu, Dong-song Bi. "Clinical risk factors for anastomotic leakage after laparoscopic anterior resection for rectal cancer: a systematic review and meta-analysis", Surgical Endoscopy, 2015

13 words — < 1%

Crossref

---

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS