ISSN 1949-8470 (online)

World Journal of *Radiology*

World J Radiol 2023 December 28; 15(12): 338-369

Published by Baishideng Publishing Group Inc

WJR

World Journal of Radiology

Contents

Monthly Volume 15 Number 12 December 28, 2023

ORIGINAL ARTICLE

Retrospective Study

Observational Study

World Journal of Radiology

Monthly Volume 15 Number 12 December 28, 2023

Contents

ABOUT COVER

Peer Reviewer of *World Journal of Radiology*, Lee K Rousslang, MD, Resident Physician, Department of Radiology, Tripler Army Medical Center, Honolulu, HI 96859, United States. lee.k.rousslang.civ@mail.mil

AIMS AND SCOPE

The primary aim of *World Journal of Radiology* (*WJR*, *World J Radiol*) is to provide scholars and readers from various fields of radiology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

WJR mainly publishes articles reporting research results and findings obtained in the field of radiology and covering a wide range of topics including state of the art information on cardiopulmonary imaging, gastrointestinal imaging, genitourinary imaging, musculoskeletal imaging, neuroradiology/head and neck imaging, nuclear medicine and molecular imaging, pediatric imaging, vascular and interventional radiology, and women's imaging.

INDEXING/ABSTRACTING

The *WJR* is now abstracted and indexed in PubMed, PubMed Central, Emerging Sources Citation Index (Web of Science), Reference Citation Analysis, China Science and Technology Journal Database, and Superstar Journals Database. The 2023 Edition of Journal Citation Reports® cites the 2022 impact factor (IF) for *WJR* as 2.5; IF without journal self cites: 2.3; 5-year IF: 2.5; Journal Citation Indicator: 0.54.

RESPONSIBLE EDITORS FOR THIS ISSUE

NAME OF JOURNAL
World Journal of Radiology

ISSN
ISSN 1949-8470 (online)

LAUNCH DATE
January 31, 2009

FREQUENCY
Monthly

EDITORS-IN-CHIEF
Thomas J Vogl

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/1949-8470/editorialboard.htm

PUBLICATION DATE
December 28, 2023

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/GerInfo/239

ONLINE SUBMISSION
https://www.f6publishing.com

World Journal of Radiology

Submit a Manuscript: https://www.f6publishing.com

World J Radiol 2023 December 28; 15(12): 338-349

DOI: 10.4329/wjr.v15.i12.338

ISSN 1949-8470 (online)

ORIGINAL ARTICLE

Retrospective Study

Deep learning-based magnetic resonance imaging reconstruction for improving the image quality of reduced-field-of-view diffusion-weighted imaging of the pancreas

Yukihisa Takayama, Keisuke Sato, Shinji Tanaka, Ryo Murayama, Nahoko Goto, Kengo Yoshimitsu

Specialty type: Radiology, nuclear medicine and medical imaging

Provenance and peer review: Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B, B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Zhao CF, China; Xiang-Ke Niu, China

Received: October 22, 2023
Peer-review started: October 22, 2023
First decision: November 2, 2023
Revised: November 12, 2023
Accepted: December 4, 2023
Article in press: December 4, 2023
Published online: December 28, 2023

Yukihisa Takayama, Keisuke Sato, Shinji Tanaka, Ryo Murayama, Nahoko Goto, Kengo Yoshimitsu, Department of Radiology, Faculty of Medicine, Fukuoka University, Fukuoka 8140180, Japan

Corresponding author: Yukihisa Takayama, MD, PhD, Associate Professor, Department of Radiology, Faculty of Medicine, Fukuoka University, 7-45-1 Nanakuma, Jonan-ku, Fukuoka 8140180, Japan. ytakayama@fukuoka-u.ac.jp

Abstract

BACKGROUND

It has been reported that deep learning-based reconstruction (DLR) can reduce image noise and artifacts, thereby improving the signal-to-noise ratio and image sharpness. However, no previous studies have evaluated the efficacy of DLR in improving image quality in reduced-field-of-view (reduced-FOV) diffusion-weighted imaging (DWI) [field-of-view optimized and constrained undistorted single-shot (FOCUS)] of the pancreas. We hypothesized that a combination of these techniques would improve DWI image quality without prolonging the scan time but would influence the apparent diffusion coefficient calculation.

AIM

To evaluate the efficacy of DLR for image quality improvement of FOCUS of the pancreas.

METHODS

This was a retrospective study evaluated 37 patients with pancreatic cystic lesions who underwent magnetic resonance imaging between August 2021 and October 2021. We evaluated three types of FOCUS examinations: FOCUS with DLR (FOCUS-DLR+), FOCUS without DLR (FOCUS-DLR−), and conventional FOCUS (FOCUS-conv). The three types of FOCUS and their apparent diffusion coefficient (ADC) maps were compared qualitatively and quantitatively.

RESULTS

FOCUS-DLR+ (3.62, average score of two radiologists) showed significantly better qualitative scores for image noise than FOCUS-DLR− (2.62) and FOCUS-conv (2.88) ($P < 0.05$). Furthermore, FOCUS-DLR+ showed the highest contrast ratio (CR) between the pancreatic parenchyma and adjacent fat tissue for b-values of 0

and 600 s/mm² (0.72 ± 0.08 and 0.68 ± 0.08) and FOCUS-DLR− showed the highest CR between cystic lesions and the pancreatic parenchyma for the b-values of 0 and 600 s/mm² (0.62 ± 0.21 and 0.62 ± 0.21) ($P < 0.05$), respectively. FOCUS-DLR+ provided significantly higher ADCs of the pancreas and lesion (1.44 ± 0.24 and 3.00 ± 0.66) compared to FOCUS-DLR− (1.39 ± 0.22 and 2.86 ± 0.61) and significantly lower ADCs compared to FOCUS-conv (1.84 ± 0.45 and 3.32 ± 0.70) ($P < 0.05$), respectively.

CONCLUSION

This study evaluated the efficacy of DLR for image quality improvement in reduced-FOV DWI of the pancreas. DLR can significantly denoise images without prolonging the scan time or decreasing the spatial resolution. The denoising level of DWI can be controlled to make the images appear more natural to the human eye. However, this study revealed that DLR did not ameliorate pancreatic distortion. Additionally, physicians should pay attention to the interpretation of ADCs after DLR application because ADCs are significantly changed by DLR.

Key Words: Deep learning-based reconstruction; Magnetic resonance imaging; Reduced field-of-view; Diffusion-weighted imaging; Pancreas

Core Tip: This study evaluated the efficacy of deep learning-based reconstruction (DLR) for image quality improvement in reduced-field-of-view diffusion-weighted imaging (DWI) of the pancreas. DLR can significantly denoise images without prolonging the scan time or decreasing the spatial resolution. The denoising level of DWI can be controlled to make the images appear more natural to the human eye. However, this study revealed that DLR did not ameliorate pancreatic distortion. Additionally, physicians should pay attention to the interpretation of apparent diffusion coefficients (ADCs) after DLR application because ADCs are significantly changed by DLR.

Citation: Takayama Y, Sato K, Tanaka S, Murayama R, Goto N, Yoshimitsu K. Deep learning-based magnetic resonance imaging reconstruction for improving the image quality of reduced-field-of-view diffusion-weighted imaging of the pancreas. *World J Radiol* 2023; 15(12): 338-349

URL: https://www.wjgnet.com/1949-8470/full/v15/i12/338.htm

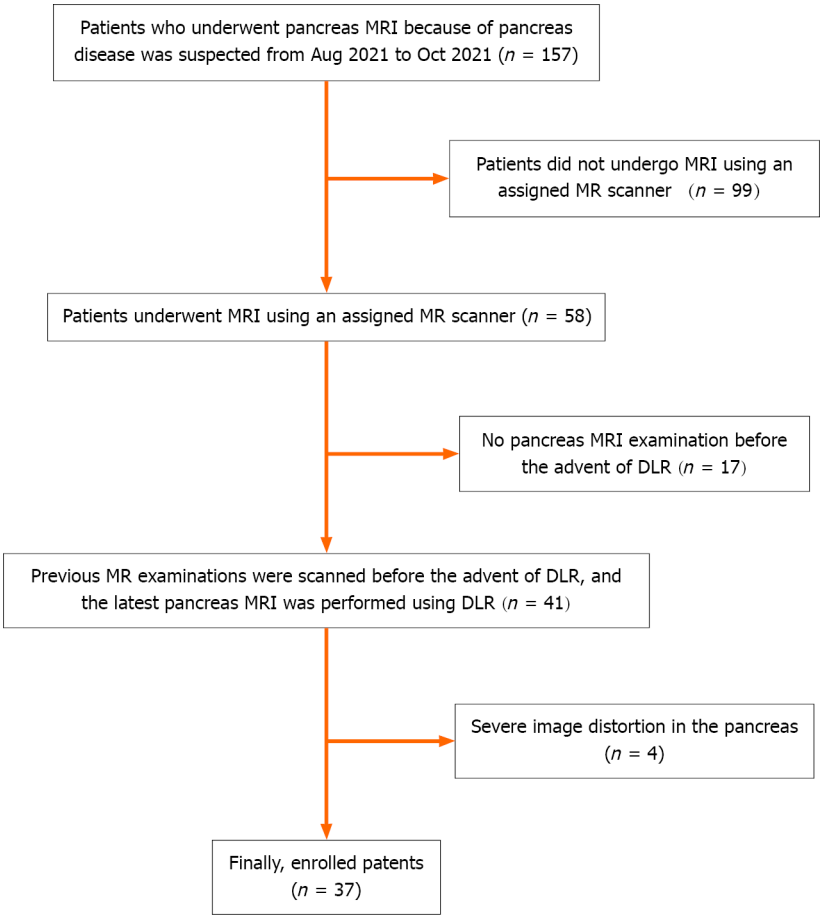
DOI: https://dx.doi.org/10.4329/wjr.v15.i12.338

INTRODUCTION

Diffusion-weighted imaging (DWI) is a widely adopted magnetic resonance imaging (MRI) technique in clinical practice [1-3]. DWI is useful for detecting and characterizing malignant and non-malignant tumors[2,4]. The detection of pancreatic cancer using DWI has been reported to be equivalent to that using dynamic contrast-enhanced computed tomography when DWI is added to magnetic resonance cholangiopancreatography (MRCP)[5,6]. DWI can be used to predict the histological grade of pancreatic neuroendocrine tumors and differentiate malignant from benign intraductal papillary neoplasms (IPMNs)[7-9].

The diagnosis of abdominal lesions based on DWI can be difficult due to artifacts such as motion, ghosting, and distortion; the pancreas is especially susceptible to these artifacts because it exists deep in the abdomen. Reduced-field-of-view (reduced-FOV) DWI is one solution to reduce artifacts in DWI[8,10-13]. In particular, imaging of the pancreas has been shown to improve image quality, such as visualization of anatomical structures, contrast-to-noise ratio (CNR), and lesion conspicuity, and reduce artifacts, such as ghosting, susceptibility, motion, and aliasing artifacts, compared to full-FOV DWI[8,10,14]. Further improvements in the image quality of pancreatic DWI would allow radiologists to detect pancreatic tumors earlier, especially small pancreatic lesions, and help predict tumor malignancy or aggressiveness.

Recently, deep learning (DL) has been applied to radiology for the detection of lesions, evaluation, and image segmentation[9,15,16]. DL is a subcategory of machine learning; therefore, a subset of artificial intelligence[17,18]. DL-based reconstruction (DLR) can reduce image noise and truncation artifacts, improving the signal-to-noise ratio (SNR) and the sharpness of anatomical structures and lesions[9,15,16]. We hypothesized that a combination of reduced-FOV DWI and DLR would improve the DWI image quality of the pancreas without prolonging scan time. To the best of our knowledge, no previous studies have evaluated this hypothesis. This study aimed to evaluate the efficacy of DLR in improving the image quality in reduced-FOV DWI of the pancreas.



DOI: 10.4329/wjr.v15.i12.338

Figure 1 A summary flowchart of the patient selection. MRI: Magnetic resonance imaging; MR: Magnetic resonance; DLR: Deep learning-based reconstruction.

MATERIALS AND METHODS

Patients

This study was approved by the Institutional Review Board of our hospital. The requirement of written informed consent was waived because this was a retrospective analysis of image post-processing of clinical magnetic resonance (MR) data. Between August 2021 and October 2021, 157 consecutive patients who underwent pancreatic MRI at our institute were investigated. The inclusion criteria were as follows: (1) Patients undergoing annual MRI studies for follow-up of pancreatic cystic lesions; (2) whose MR examinations were performed using an assigned MR scanner; and (3) previous pancreatic MR images were scanned before the advent of DLR, but the latest pancreatic MR images were performed using DLR. Patients were excluded if the pancreas could not be evaluated due to severe distortion. Ultimately, 37 patients [15 females, 22 males; median age (range), 66 years (41–85 years)] were enrolled. Figure 1 shows a flow chart of the patient selection process. Among the 37 patients, 21 were suspected to have IPMN on endoscopic retrograde cholangiopancreatography and/or MRI. The other 16 patients were diagnosed with unspecified pancreatic cystic lesions, such as IPMN and lymphoepithelial cysts, but their diagnosis was not confirmed.

MRI

All examinations were performed using a clinical 3.0-Tesla MR system (Discovery MR750w 3.0T; GE Healthcare, Waukesha, WI, United States). In addition to routine MRI, such as T1-weighted imaging, T2-weighted imaging, full-FOV DWI, and MRCP, each patient underwent reduced-FOV DWI with field-of-view optimized and constrained undistorted single-shot (FOCUS).

The DLR, *i.e.*, AIR™ Recon DL (GE Healthcare), is a vendor-supplied MRI reconstruction algorithm based on a deep convolutional network trained on a database of more than 10000 pairs of artifact-free, high-SNR, high-spatial-resolution image, plus the corresponding low SNR, low-spatial-resolution images[19]. It converts truncation artifacts into improved image sharpness while simultaneously denoising the images[19]. The AIR™ Recon DL was already trained before being installed on the assigned MRI machine, so it was ready to integrate into our MRI reconstruction pipeline. Our motivation for introducing DLR was to improve the image quality of FOCUS of the pancreas, because it suffers from a low SNR and the limitation of not providing good results at higher b-value settings.

Table 1 Details of imaging parameters

Parameters	FOCUS-DLR+/–	FOCUS-conv
Repetition time, ms	3000–10000	3500–15000
Echo time, ms	60	60
Flip angle, degree	90	90
FOV, mm^2	220 × 110	220 × 110
Matrix	120 × 64	130 × 40
FOV reduction	Anterior-posterior	Anterior-posterior
Slice thickness, mm	3	4
Slice gap, mm	3	5
Number of slices	20–30	15–20
Number of excitations	4	8
b-values, s/mm^2	0 and 600	0 and 600
Band width, Hz/pixel	1950	1300
Respiratory compensation	Respiratory-triggered with navigator echo	Respiratory-triggered with or without navigator echo
Deep learning reconstruction factor	Moderate	N/A
Scan time, min	2–5	3–10

The repetition time, number of slices and scan time varied depending on the patients' condition. DLR: Deep learning-based reconstruction; FOCUS: Field-of-view optimized and constrained undistorted single-shot; FOCUS-conv: Conventional field-of-view optimized and constrained undistorted single-shot; FOV: Field of view; N/A: Not applicable.

In this study, three types of FOCUS were evaluated. First, two types of FOCUS images are generated from a single set of raw FOCUS scanning data. One type of FOCUS was reconstructed with the use of DLR; it is referred to hereafter as "FOCUS-DLR+." The other type of FOCUS was reconstructed without DLR ("FOCUS-DLR−"). Furthermore, all enrolled patients had undergone a previous MR examination that included a conventional FOCUS ("FOCUS-conv"), which was widely used in clinical practice before the advent of DLR, and before other improvements that are currently standard on the MR scanner. The average difference (range) in the length of time between FOCUS-DLR+/− and FOCUS-conv was 842.8 (181–2007) d.

The details of the imaging parameters of FOCUS-DLR+/− and FOCUS-conv are shown in Table 1. A two-dimensional (2D) spatially selective echoplanar radiofrequency (RF) excitation pulse was used for FOCUS. This reduces the excitation volume in the phase-encoding and slice-selective directions[11]. In a 2D RF pulse, the displacement between fat and water is designed such that the excited fat profile is completely outside the excited water profile; therefore, a fat-suppression technique is unnecessary[11]. A b-value of 600 s/mm^2 was used as the maximum b-value in this study, because FOCUS-conv with a b-value of 600 s/mm^2 provided acceptable image quality to visualize the pancreatic parenchyma. We also obtained an apparent diffusion coefficient (ADC) map for each type of FOCUS based on the signal intensity (SI) decay of each pixel on DWI with b-values of 0 and 600 s/mm^2.

Image assessment

We conducted qualitative and quantitative comparisons among the three FOCUS types and their ADC maps. The comparison between FOCUS-DLR+ and FOCUS-DLR− was aimed at evaluating DLR by comparing the efficacy of DLR to improve DWI image quality and assessing ADC maps. We also conducted a comparison between FOCUS-DLR+ and FOCUS-conv and between FOCUS-DLR− and FOCUS-conv. This was because the differences between FOCUS-DLR+/− and FOCUS-conv included not only the use of DLR but also updates to the MR scanner, including the update of the MR console to include the AIR™ Recon software.

A study coordinator (Takayama Y, with 23 years of experience in interpreting abdominal MRI) searched and displayed the patients' MRI datasets using a picture archiving and communication system (PACS) (Rapideye, Canon Medical Systems, Tokyo). For qualitative comparison: (1) Sharpness of the pancreatic contour; (2) image noise; (3) distortion of the pancreas; (4) visualization of pancreatic cystic lesions; and (5) visualization of the main pancreatic duct (MPD) were independently evaluated by two radiologists (R1: Tanaka S and R2: Sato K, with 7 and 6 years of experience in interpreting abdominal MRI, respectively), who were blinded to imaging information and the patient's clinical data. (1), (2), (3), and (4) were evaluated using each type of FOCUS with a b-value of 600 s/mm^2, and (5) was evaluated using ADC maps. Qualitative assessments were performed using a 4-point scoring system. The image-quality scores are listed in Table 2.

Table 2 Image-quality scores in the qualitative assessment

Qualitative assessment	1	2	3	4
Sharpness of the pancreas contour	Entire pancreas contour is unclear	< 50% of the pancreas contour is clear	≥ 50% of the pancreas contour is clear	Entire pancreas contour is clear
Image noise	Severe noise; no visualization of any organs	Moderate noise; compromised diagnostic capability of FOCUS is more than ≥ 50% of the image	Mild noise; compromised the diagnostic capability of FOCUS is < 50% of the image	No or slight noise on the image
Distortion of the pancreas	Severe distortion; no visualization of the whole pancreas	Moderate distortion; no visualization is ≥ 50% of the pancreas	Mild distortion; no visualization is < 50% of the pancreas	No distortion of the pancreas
Visualization of pancreas cystic lesion	No visualization of pancreas cystic lesion	Pancreas cystic lesion is visible, but its SI is low	Pancreas cystic lesion is visible with high SI, but its contour is unclear	Pancreas cystic lesion is clearly visible with high SI and clear contour
Visualization of MPD on ADC map	No visualization of MPD	Visible MPD is < 50% of the pancreas	Visible MPD is ≥ 50% of the pancreas	Whole MPD is visible

MPD: Main pancreatic duct; ADC: Apparent diffusion coefficient; SI: Signal intensity; FOCUS: Field-of-view optimized and constrained undistorted single-shot.

For quantitative comparison, we calculated the following: (1) Contrast ratios (CRs) between the pancreatic parenchyma and adjacent fat tissue, hereafter referred to as "$CR_{pancreas\text{-}fat}$"; (2) CRs between pancreatic cystic lesion and pancreatic parenchyma ("$CR_{lesion\text{-}pancreas}$"); (3) the ADC of pancreatic parenchyma ("$ADC_{pancreas}$"); and (4) the ADC of pancreatic cystic lesions ("ADC_{lesion}") for the three types of FOCUS and their ADC maps after drawing polygonal regions of interest (ROIs) on the pancreatic parenchyma, adjacent fat tissue and pancreatic cystic lesion. CRs were calculated for each type of FOCUS, with b-values of 0 and 600 s/mm^2. CRs and ADCs were evaluated by the same two radiologists (R1 and R2) using the same PACS after completion of the qualitative assessments. $CR_{pancreas\text{-}fat}$ and $CR_{lesion\text{-}pancreas}$ were calculated instead of SNR or CNR because: (1) FOCUS does not include background air within the imaging area; and (2) a parallel imaging technique was used for the scan; thus, it was impossible to measure background air noise. CRs were calculated using the following formula:

$CR_{pancreas\text{-}fat} = (SI_{pancreas} - SI_{fat})/ (SI_{panceras} + SI_{fat})$

$CR_{lesion\text{-}pancreas} = (SI_{lesion} - SI_{pancreas})/ (SI_{lesion} + SI_{pancreas})$

$SI_{pancreas}$ is the SI of the pancreatic parenchyma, SI_{fat} is the SI of adjacent fat tissue, and SI_{lesion} is the SI of the pancreatic cystic lesion. In this study, all CRs are presented as absolute values.

Regarding the calculation of CRs and ADCs, the routine MRI findings of the patients were used for the localization of MPD and pancreatic cystic lesions. To calculate $CR_{pancreas\text{-}fat}$ and $ADC_{pancreas}$, three as-large-as-possible polygonal ROIs were drawn for each patient on the head, body, and tail of the pancreas to avoid MPD, lesions, and artifacts on the FOCUS images using b-values of 0 and 600 s/mm^2. Three additional large-as-possible polygonal ROIs were drawn near the head, body, and tail of the pancreas to avoid vessels, lesions, air, and artifacts for each patient. For $CR_{lesion\text{-}pancreas}$ and ADC_{lesion}, as-large-as-possible polygonal ROIs were drawn within the pancreatic cystic lesion and the adjacent pancreatic parenchyma on the same axial slice where the lesions showed the maximum diameter. If there were several lesions in the pancreas, the largest lesion was selected for the calculation.

The same ROIs were duplicated for the FOCUS-DLR+, FOCUS-DLR−, and ADC maps. ROIs of similar size for FOCUS-conv and its ADC map were drawn as those of FOCUS-DLR+/−. In addition to qualitative and quantitative comparisons, we compared the scan time between FOCUS-DLR+/− and FOCUS-conv.

Statistical analysis

To compare image-quality scores, CRs, and ADCs among the three types of FOCUS, the Friedman test was performed. When the Friedman test showed a significant result, the Bonferroni post-hoc test was performed for pairwise comparisons among the three types of FOCUS.

The inter-reader agreement between the image-quality scores of the two radiologists was analyzed using weighted kappa statistics. The kappa values are interpreted as follows: < 0: No agreement; 0–0.20: Slight agreement; 0.21–0.40: Fair agreement; 0.41–0.60: Moderate agreement; 0.61–0.80: Substantial agreement; and 0.81–1.00: Almost perfect agreement.

Comparisons of CRs and ADCs among the three types of FOCUS were analyzed after the measurement results of the two radiologists were combined because it was difficult for them to draw the same ROIs at the same locations of the pancreatic parenchyma, adjacent fat tissue, and pancreatic cystic lesions. Finally, the paired t-test was performed for the comparison of scan time between FOCUS-DLR+/− and FOCUS-conv. All statistical analyses were performed with IBM SPSS Statistics 25.0 (IBM Japan, Tokyo). For the Friedman test and paired t-test, P values <0.05 were considered significant and $P < 0.0167$ (0.05/3) for the Bonferroni post-hoc test was considered significant.

Table 3 The results of the two radiologists' qualitative assessments

Qualitative assessment	Reader	Mean image-quality score			*P* value			
					Friedman test	Bonferroni post-hoc test		
		FOCUS-DLR+	FOCUS-DLR-	FOCUS-conv		FOCUS-DLR+ *vs* FOCUS-DLR-	FOCUS-DLR+ *vs* FOCUS-conv	FOCUS-DLR- *vs* FOCUS-conv
Sharpness of pancreas contour	R1	3.32	3	2.24	< 0.001[1]	0.31	< 0.001[1]	0.0011
	R2	3.32	3.03	2.05	< 0.001[1]	0.49	< 0.001[1]	< 0.001[1]
Image noise	R1	3.65	2.7	2.86	< 0.00 1[1]	< 0.001[1]	< 0.001[1]	1
	R2	3.59	2.54	2.7	< 0.001[1]	< 0.001[1]	< 0.001[1]	0.97
Distortion of pancreas	R1	3.16	3.11	3.05	0.05	N/A	N/A	N/A
	R2	3.11	3.18	3.05	0.73	N/A	N/A	N/A
Visualization of pancreas cystic lesion	R1	3.7	3.49	2.89	< 0.001[1]	0.67	0.0161	0.35
	R2	3.62	3.32	2.7	< 0.001[1]	0.24	0.0011	0.17
Visualization of MPD on ADC map	R1	2.97	2.54	2.19	< 0.001[1]	0.17	0.0031	0.49
	R2	2.73	2.24	1.99	< 0.001[1]	0.11	0.0011	0.35

[1]A significant difference.

DLR: Deep learning-based reconstruction; FOCUS: Field-of-view optimized and constrained undistorted single-shot; FOCUS-conv: Conventional field-of-view optimized and constrained undistorted single-shot; ADC: Apparent diffusion coefficient; MPD: Main pancreatic duct; N/A: Not available.

RESULTS

Qualitative image assessments

The detailed results of the qualitative assessments by the two radiologists are shown in Table 3. Radiologists obtained similar results. The Friedman test showed significant differences between the three types of FOCUS in image-quality scores for pancreatic contour sharpness, image noise, visualization of pancreatic cystic lesions, and visualization of MPD on the ADC map ($P < 0.05$). There were no significant differences in the image-quality scores for pancreatic distortion among the three types of FOCUS ($P > 0.05$).

The Bonferroni post-hoc test revealed that FOCUS-DLR+ and FOCUS-DLR− showed significantly higher image-quality scores for the sharpness of the pancreas contour than FOCUS-conv ($P < 0.0167$), but there were no significant differences between FOCUS-DLR+ and FOCUS-DLR− ($P > 0.0167$). Regarding image-quality scores of the image noise, FOCUS-DLR+ showed significantly higher scores than FOCUS-DLR− and FOCUS-conv ($P < 0.0167$), but there were no significant differences in scores between FOCUS-DLR− and FOCUS-conv ($P > 0.0167$). FOCUS-DLR+ showed significantly higher image-quality scores for visualization of the pancreatic cystic lesion and visualization of MPD on the ADC map compared to FOCUS-conv ($P < 0.0167$), but there were no significant differences between FOCUS-DLR+ and FOCUS-DLR− or between FOCUS-DLR+ and FOCUS-conv ($P > 0.0167$).

Inter-reader agreements

Table 4 provides the results of the inter-reader agreement between the two radiologists. All qualitative assessments showed significant agreement ($P < 0.001$).

Quantitative image assessments

The average (range) of the ROIs of the pancreatic parenchyma, the adjacent fat tissue, and the cystic lesion of the pancreas drawn by the two radiologists were the following: R1, 170.9 mm² (57.3–325.0 mm²), 235.3 mm² (50.2–791.75 mm²) and 92.7 mm² (22.2–457.6 mm²); R2, 236.3 mm² (72.4–676.53 mm²), 151.5 ± 38.2 mm² (70.9–253.66 mm²) and 104.6 mm² (22.2–551.0 mm²), respectively.

The detailed results of the quantitative assessment are presented in Table 5. The Friedman test showed significant differences between the three types of FOCUS regarding $CR_{pancreas\text{-}fat}$ using b-values of 0 and 600 s/mm², $CR_{lesion\text{-}pancreas}$ using b-values of 0 and 600 s/mm², $ADC_{pancreas}$ and ADC_{lesion} ($P < 0.05$).

The Bonferroni post-hoc test revealed that FOCUS-DLR+ showed significantly higher $CR_{pancreas\text{-}fat}$ using b-values of 0 and 600 s/mm² compared to FOCUS-DLR− and FOCUS-conv, and FOCUS-DLR− showed significantly higher $CR_{pancreas\text{-}fat}$ than FOCUS-conv ($P < 0.0167$).

FOCUS- DLR− showed a significantly higher $CR_{lesion\text{-}pancreas}$ using b-values of 0 and 600 s/mm² compared to FOCUS-DLR+ and FOCUS-conv, and FOCUS-DLR+ showed significantly higher $CR_{lesion\text{-}pancreas}$ than FOCUS-conv ($P < 0.0167$).

Table 4 Inter-reader agreement between the two radiologists

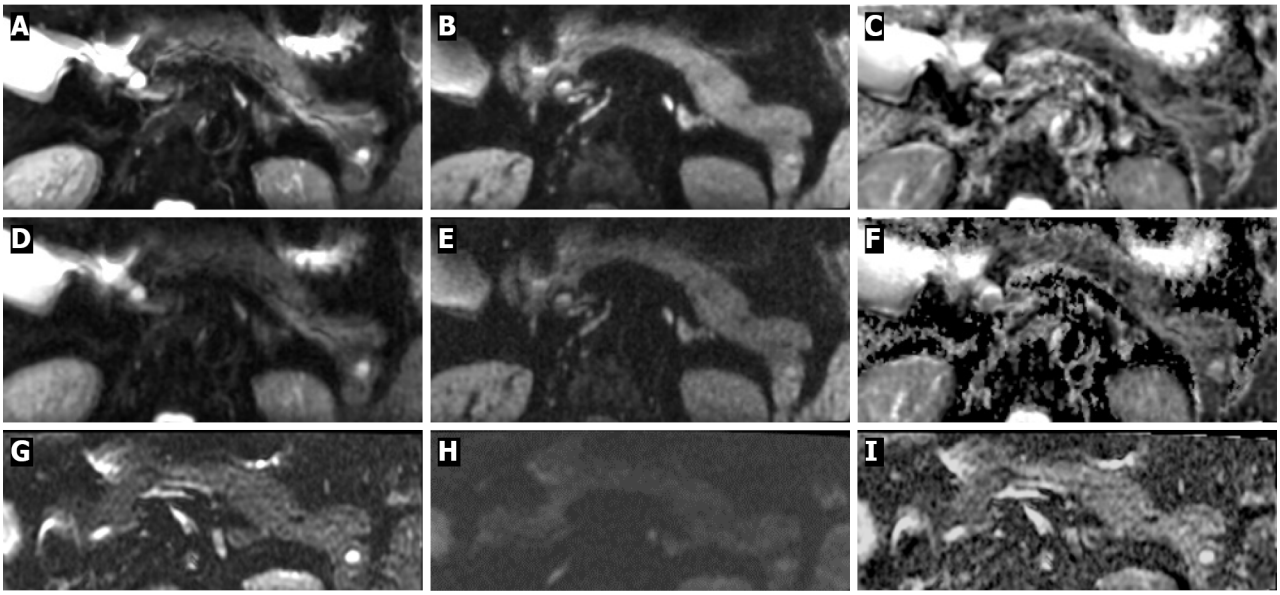
Qualitative assessment	Imaging	κ value (95%CI)	*P* value
Sharpness of pancreas contour	FOCUS-DLR+	0.73 (0.52–0.95)	< 0.001[1]
	FOCUS-DLR-	0.69 (0.45–0.93)	< 0.001[1]
	FOCUS-conv	0.66 (0.43–0.89)	< 0.001[1]
Image noise	FOCUS-DLR+	0.69 (0.46–0.92)	< 0.001[1]
	FOCUS-DLR-	0.70 (0.50–0.90)	< 0.001[1]
	FOCUS-conv	0.64 (0.42–0.87)	< 0.001[1]
Distortion of pancreas	FOCUS-DLR+	0.61 (0.40–0.82)	< 0.001[1]
	FOCUS-DLR-	0.71 (0.53–0.89)	< 0.001[1]
	FOCUS-conv	0.61 (0.41–0.82)	< 0.001[1]
Visualization of pancreas cystic lesion	FOCUS-DLR+	0.80 (0.56–1.03)	< 0.001[1]
	FOCUS-DLR-	0.76 (0.60–0.93)	< 0.001[1]
	FOCUS-conv	0.71 (0.51–0.91)	< 0.001[1]
Visualization of MPD on ADC map	FOCUS-DLR+	0.67 (0.51–0.83)	< 0.001[1]
	FOCUS-DLR-	0.75 (0.60–0.90)	< 0.001[1]
	FOCUS-conv	0.74 (0.54–0.94)	< 0.001[1]

[1]A significant difference. *P* values were obtained by the weighted κ statistic.
κ values: < 0, no agreement; 0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement. 95%CI: 95% confidence intervals; DLR: Deep learning-based reconstruction; FOCUS: Field-of-view optimized and constrained undistorted single-shot; FOCUS-conv: Conventional field-of-view optimized and constrained undistorted single-shot; ADC: Apparent diffusion coefficient; MPD: Main pancreatic duct.

Table 5 The results of quantitative assessments

Quantitative assessments	mean ± SD			*P* values			
				Friedman test	Bonferroni post-hoc test		
	FOCUS-DLR+	FOCUS-DLR-	FOCUS-conv		FOCUS-DLR+ *vs* FOCUS-DLR-	FOCUS-DLR+ *vs* FOCUS-conv	FOCUS-DLR- *vs* FOCUS-conv
$CR_{pancreas\text{-}fat}$ on FOCUS with b value of 600 s/mm^2	0.68 ± 0.08	0.49 ± 0.10	0.27 ± 0.11	< 0.001[1]	< 0.001[1]	< 0.001[1]	< 0.001[1]
$CR_{pancreas\text{-}fat}$ on FOCUS with b value of 0 s/mm^2	0.72 ± 0.08	0.65 ± 0.08	0.40 ± 0.11	< 0.001[1]	< 0.001[1]	< 0.001[1]	< 0.001[1]
$CR_{lesion\text{-}pancreas}$ on FOCUS with b value of 600 s/mm^2	0.51 ± 0.26	0.62 ± 0.21	0.01 ± 0.26	< 0.001[1]	< 0.001[1]	< 0.001[1]	< 0.001[1]
$CR_{lesion\text{-}pancreas}$ on FOCUS with b value of 0 s/mm^2	0.53 ± 0.21	0.62 ± 0.21	0.17 ± 0.19	< 0.001[1]	< 0.001[1]	< 0.001[1]	< 0.001[1]
$ADC_{pancreas+}$ (× 10^{-3} mm^2/s)	1.44 ± 0.24	1.39 ± 0.22	1.84 ± 0.45	< 0.001[1]	< 0.001[1]	< 0.001[1]	< 0.001[1]
ADC_{lesion} (× 10^{-3} mm^2/s)	3.00 ± 0.66	2.86 ± 0.61	3.32 ± 0.70	< 0.001[1]	< 0.001[1]	< 0.001[1]	< 0.001[1]

[1]A significant difference.
mean ± SD were calculated from the combined data of two radiologists' measurements. All CRs are presented in absolute values. 95%CI: 95% confidence intervals; DLR: Deep learning-based reconstruction; FOCUS: Field-of-view optimized and constrained undistorted single-shot; FOCUS-conv: Conventional field-of-view optimized and constrained undistorted single-shot; ADC: Apparent diffusion coefficient; MPD: Main pancreatic duct; CR: Contrast ratio.



DOI: 10.4329/wjr.v15.i12.338

Figure 2 A 45-year-old male with an unspecified pancreas cyst. Reduced-field-of-view diffusion-weighted image [field-of-view optimized and constrained undistorted single-shot (FOCUS)] with b-values of 0 and 600 s/mm^2, and apparent diffusion coefficient (ADC) maps of FOCUS-deep learning-based reconstruction (DLR)+ (A-C), FOCUS-DLR− (D-F), and conventional FOCUS (FOCUS-conv) (G-I). FOCUS-DLR+ provides reduced image noise and better contrast ratio between the pancreas parenchyma and adjacent fat tissue [contrast ratios (CR)$_{pancreas-fat}$] compared to FOCUS-DLR- and FOCUS-conv. Both FOCUS-DLR+ and FOCUS-DLR- show better image-quality scores for sharpness of pancreas contour, CR$_{pancreas-fat}$, and contrast ratio between the pancreas cystic lesion and pancreas parenchyma (CR$_{lesion-pancreas}$) compared to FOCUS-conv. FOCUS-DLR- provides a better CR$_{lesion-pancreas}$ than FOCUS-DLR+, but there is no difference in the image-quality score for visualization of the cystic lesion of the pancreas between FOCUS-DLR+ and FOCUS-DLR-. The average image-quality scores from two radiologists for FOCUS-DLR+, FOCUS-DLR- and FOCUS-conv are as follows: Sharpness of the pancreas contour, 3.5, 3.0, and 2.0; image noise, 3.5, 2.5, and 2.0; and visualization of pancreatic cystic lesion, 3.0, 2.5, 1.0, respectively. The average CR$_{pancreas-fat}$ and CR$_{lesion-pancreas}$ on FOCUS with a b-value of 600 s/mm^2 from two radiologists for FOCUS-DLR+, FOCUS-DLR- and FOCUS-conv are 0.54 and 0.75, 0.38 and 0.81, and 0.23 and 0.10, respectively. The average ADCs of the pancreatic parenchyma and pancreatic cystic lesion of two radiologists for FOCUS-DLR+, FOCUS-DLR- and FOCUS-conv are 1.59 and 2.28, 1.49 and 2.26, and 1.93 and 4.01 (× 10^{-3} mm^2/s), respectively.

FOCUS-conv showed significantly higher ADC$_{pancreas}$ and ADC$_{lesion}$ compared to FOCUS-DLR+ and FOCUS-DLR−, and FOCUS-DLR+ showed significantly higher ADC$_{pancreas}$ and ADC$_{lesion}$ compared to FOCUS-DLR− ($P < 0.0167$).

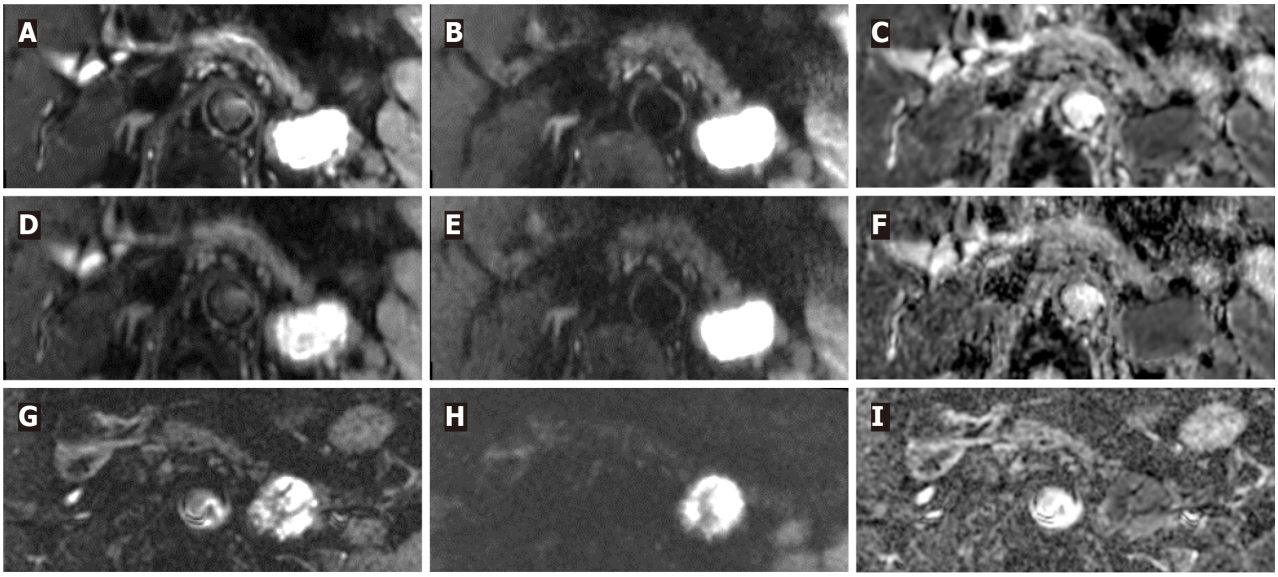
The average scan time of FOCUS-DLR+/− (3 min 27 s) was significantly shorter than that of FOCUS-conv (6 min 28 S) ($P < 0.001$). Figures 2 and 3 show representative images of FOCUS-DLR+, FOCUS-DLR−, and FOCUS-conv.

DISCUSSION

Our findings showed that FOCUS-DLR+ can significantly denoise images without prolonging the scan time or decreasing the spatial resolution compared to FOCUS-DLR− and FOCUS-conv. This result is consistent with studies that analyzed the effectiveness of DLR in brain, musculoskeletal, and prostate MRI examinations[8,16,20]. DLR has demonstrated superiority over other denoising methods. Filter-based noise reduction is commonly applied to data reconstruction pipelines to mitigate image noise[21]. However, this method removes image noise and degrades SIs of structural details, resulting in blurred images[9]. On average, an increased number of signals is also effective in obtaining higher-SNR images; however, this method requires longer scan times[4]. A decrease in spatial resolution can reduce image noise because the image SNR is proportional to the voxel size[11]. However, the decrease in spatial resolution is a disadvantage for diagnosis, especially for the depiction of small lesions.

Another benefit of DLR for denoising is that it can control the level of denoising of DWI to make the images appear more natural to the human eye. DLR can improve CR$_{pancreas-fat}$ on FOCUS using a b-value of 600 s/mm^2, and CR$_{pancreas-fat}$ and CR$_{lesion-pancreas}$ on FOCUS using a b-value of 0 s/mm^2. We speculated that a higher CR would clarify the pancreatic parenchyma and lesions. In fact, FOCUS-DLR− showed higher CR$_{lesion-pancreas}$ than FOCUS-DLR+ with a b-value of 600 s/mm^2. This result could be related to an increased in SIs of the pancreatic parenchyma on FOCUS-DLR+ compared to that on FOCUS-DLR−. However, the results of CR$_{lesion-pancreas}$ on FOCUS with a b-value of 600 s/mm^2 did not indicate that the detection of pancreatic cystic lesions would be affected by the use of DLR. Instead, DLR is helpful to determine whether there is a lesion inside or outside the tissue.

FOCUS-DLR+ showed a higher image-quality score for the sharpness of the pancreas contour compared to FOCUS-conv, but no significant differences were observed between FOCUS-DLR+ and FOCUS-DLR−. We suggest that FOCUS-DLR+ may be effective in visualizing anatomical structures and lesions in the pancreas. DLR has been reported to be useful in improving image sharpness because it can effectively eliminate truncation artifacts, while denoising is



DOI: 10.4329/wjr.v15.i12.338 **Copyright** ©The Author(s) 2023.

Figure 3 A 70-year-old male with an unspecified pancreas cyst. Reduced-field-of-view diffusion-weighted image [field-of-view optimized and constrained undistorted single-shot (FOCUS)] with b-values of 0 and 600 s/mm^2, and apparent diffusion coefficient (ADC) maps of FOCUS-deep learning-based reconstruction (DLR)+ (A-C), FOCUS-DLR- (D-F), and conventional FOCUS (FOCUS-conv) (G-I). FOCUS-DLR+ provides reduced image noise and better contrast ratio between the pancreas parenchyma and adjacent fat tissue [contrast ratios (CR)$_{pancreas-fat}$] compared to FOCUS-DLR- and FOCUS-conv. Both FOCUS-DLR+ and FOCUS-DLR- show better image-quality scores for sharpness of pancreas contour, CR$_{pancreas-fat}$, and contrast ratio between the pancreas cystic lesion and pancreas parenchyma (CR$_{lesion-pancreas}$) compared to FOCUS-conv. FOCUS-DLR- provides a better CR$_{lesion-pancreas}$ than FOCUS-DLR+, but there is no difference in the image-quality score for visualization of the cystic lesion of the pancreas between FOCUS-DLR+ and FOCUS-DLR-. The average image-quality scores from two radiologists for FOCUS-DLR+, FOCUS-DLR- and FOCUS-conv are as follows: Sharpness of the pancreas contour, 4.0, 3.0, and 2.0; image noise, 4.0, 3.0, and 3.0; and visualization of pancreatic cystic lesion, 4.0, 4.0, 3.0, respectively. The average CR$_{pancreas-fat}$ and CR$_{lesion-pancreas}$ on FOCUS with a b-value of 600 s/mm^2 from two radiologists for FOCUS-DLR+, FOCUS-DLR- and FOCUS-conv are 0.54 and 0.75, 0.38 and 0.81, and 0.23 and 0.10, respectively. The average ADCs of the pancreatic parenchyma and pancreatic cystic lesion of two radiologists for FOCUS-DLR+, FOCUS-DLR- and FOCUS-conv are 1.24 and 1.15, 1.20 and 1.15, and 1.98 and 1.60 (× 10^{-3} mm^2/s), respectively.

controlled independently[9,16]. Truncation artifacts are caused by incomplete sampling of high spatial frequencies in the Fourier domain (k-space), creating edge ringing in the final reconstructed image, which can be mitigated by increasing the spatial resolution[11,16]. Reduced-FOV DWI can provide a higher spatial resolution than full-FOV DWI[11]. Thus, reduced-FOV DWI may decrease truncation artifacts, regardless of the application of DLR. We speculated that the improvement in the sharpness of the pancreatic contour could be so subtle that it could be difficult for the human eye to recognize.

Our results also revealed that DLR did not ameliorate pancreatic distortion. No previous study has concluded that DLR could be effective in improving image distortion; therefore, our current findings seem reasonable. We used the single-shot echoplanar imaging sequence, which is occasionally disturbed by distortion artifacts in the phase-encoding direction[8,10,14]. In the present study, the pancreas of some patients was distorted due to adjacent air in the gastrointestinal tract. We concluded that DLR could not modify the severe distortions of pancreatic images in the post-processing pipeline after the scan. To reduce image distortion, air must be removed within the scan area or the parameter settings must be modified.

Regarding the comparison of ADCs, FOCUS-DLR+ showed higher image-quality scores for the visualization of MPD on an ADC map compared to FOCUS-conv, but no significant differences were observed between FOCUS-DLR+ and FOCUS-DLR− or between FOCUS- DLR− and FOCUS-conv. This result may be related to the differences in image noise and CRs among the three types of FOCUS. Generally, MPD shows a higher ADC than the pancreatic parenchyma. The high ADC of the MPD is easy for the human eye to recognize on the ADC map; therefore, the DLR might not influence the qualitative assessment of the visualization of MPD on ADC maps.

ADC$_{pancreas}$ and ADC$_{lesion}$ acquired from FOCUS-DLR+ were significantly higher than those of FOCUS-DLR− and significantly lower than those of FOCUS-conv. ADCs can vary depending on the MRI apparatus, selection of b-values, and the existence of artifacts[13,22]. Image noise on DWI may also affect the calculation of ADC[23]. The ADC metrics derived from reduced-FOV DWI are controversial; both increased and decreased ADCs of reduced-FOV DWI have been reported compared to full-FOV DWI[13]. Our results indicated that ADCs could vary with the use of DLR due to differences in the SIs of the pancreatic parenchyma and pancreatic cystic lesions and in the level of image noise between the three types of FOCUS. Although ADC measurements may be helpful in differentiating malignancy from non-malignancy as a supplement to other imaging modalities, the interpretation of ADCs after DLR requires further study. One limitation of this study is that we were unable to evaluate ADC$_{pancreas}$ or ADC$_{lesion}$ by referring to standard references of pathological findings or larger patient populations. In summary, we could not estimate how DLR affected the calculation results of ADCs and lesion characterization.

There are some limitations to this study. First, we analyzed a small number of patients from a single center. It may be difficult to avoid bias in our results and speculations. A large-scale, multicenter study would be necessary to validate our results. The retrospective design of this study is also a potential source of bias. Second, none of the patients enrolled had solid pancreatic tumors, such as pancreatic carcinoma or neuroendocrine tumors. The diagnosis of such lesions on the FOCUS and ADC maps is of great interest to radiologists. Unfortunately, it was impossible to evaluate these lesions by comparing FOCUS-DLR+/− with FOCUS-conv simply because no patients with such lesions presented for annual follow-up MR examinations. The mechanism by which DLR affects lesion detectability, especially in small pancreatic carcinomas, remains unknown. Third, we used the vendor-supplied DLR that was already trained before being installed on MRI machine. On the other hand, the machine learning model is widely regarded as a black box. It meant that we could not know detailed processes of DLR to improve the image quality of FOCUS-DLR+. Although we evaluated our data using common analysis methods, it might be necessary to prove whether or not our methodology was appropriate to evaluate the effectiveness of DLR. Finally, the b-values used in the analyses of the reduced-FOV were 0 and 600 s/mm^2 for the aforementioned reasons. These b-values make it impossible to compare our findings with those of previous studies.

CONCLUSION

The use of DLR improved the image noise and CRs on FOCUS without prolonging the scan time. However, the interpretation of ADCs on FOCUS, with or without DLR, requires further study.

ARTICLE HIGHLIGHTS

Research background

A combination of these techniques would improve diffusion-weighted imaging (DWI) image quality without prolonging the scan time but would influence the apparent diffusion coefficient calculation.

Research motivation

The image quality of reduced-field-of-view DWI [field-of-view optimized and constrained undistorted single-shot (FOCUS)] of the pancreas suffers from a low signal-to-noise ratio and the limitation of not providing good results at higher b-value settings.

Research objectives

This study aimed to evaluate the efficacy of deep learning-based reconstruction (DLR) for image quality improvement of FOCUS of the pancreas.

Research methods

This was a retrospective study evaluated 37 patients with pancreatic cystic lesions who underwent magnetic resonance imaging between August 2021 and October 2021. We evaluated three types of FOCUS examinations: FOCUS with DLR (FOCUS-DLR+), FOCUS without DLR (FOCUS-DLR−), and conventional FOCUS (FOCUS-conv). The three types of FOCUS and their apparent diffusion coefficient (ADC) maps were compared qualitatively and quantitatively.

Research results

FOCUS-DLR+ (3.62, average score of two radiologists) showed significantly better qualitative scores for image noise than FOCUS-DLR− (2.62) and FOCUS-conv (2.88) ($P < 0.05$). Furthermore, FOCUS-DLR+ showed the highest contrast ratios (CRs) between the pancreatic parenchyma and adjacent fat tissue for b-values of 0 and 600 s/mm^2 (0.72 ± 0.08 and 0.68 ± 0.08) and FOCUS-DLR− showed the highest CR between cystic lesions and the pancreatic parenchyma for the b-values of 0 and 600 s/mm^2 (0.62 ± 0.21, and 0.62 ± 0.21) ($P < 0.05$), respectively. FOCUS-DLR+ provided significantly higher ADCs of the pancreas and lesion (1.44 ± 0.24 and 3.00 ± 0.66) compared to FOCUS-DLR− (1.39 ± 0.22 and 2.86 ± 0.61) and significantly lower ADCs compared to FOCUS-conv (1.84 ± 0.45 and 3.32 ± 0.70) ($P < 0.05$), respectively.

Research conclusions

DLR improved image noise and CRs on FOCUS without prolonging the scan time. However, caution should be exercised when interpreting the ADCs after DLR.

Research perspectives

This study revealed that DLR can significantly denoise images without prolonging the scan time or decreasing the spatial resolution. However, DLR did not ameliorate pancreatic distortion and physicians should pay attention to the interpretation of ADCs after DLR application.

FOOTNOTES

Author contributions: Takayama Y and Yoshimitsu K designed the research and wrote the paper; Sato K, Tanaka S, Murayama R, and Goto N contributed to data collection, data analysis, and all authors approved the final manuscript.

Institutional review board statement: This study was reviewed and approved by the Ethics Committee of Fukuoka University [Approval No. U21-966].

Informed consent statement: The requirement for written informed consent was waived because this was a retrospective analysis of image post-processing of clinical MR data.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Data sharing statement: No additional data is available.

Country/Territory of origin: Japan

ORCID number: Yukihisa Takayama 0000-0002-2029-9316.

S-Editor: Lin C
L-Editor: A
P-Editor: Lin C

REFERENCES

1 **Kwee TC**, Takahara T, Ochiai R, Nievelstein RA, Luijten PR. Diffusion-weighted whole-body imaging with background body signal suppression (DWIBS): features and potential applications in oncology. *Eur Radiol* 2008; **18**: 1937-1952 [PMID: 18446344 DOI: 10.1007/s00330-008-0968-z]

2 **Qayyum A**. Diffusion-weighted imaging in the abdomen and pelvis: concepts and applications. *Radiographics* 2009; **29**: 1797-1810 [PMID: 19959522 DOI: 10.1148/rg.296095521]

3 **Takahara T**, Imai Y, Yamashita T, Yasuda S, Nasu S, Van Cauteren M. Diffusion weighted whole body imaging with background body signal suppression (DWIBS): technical improvement using free breathing, STIR and high resolution 3D display. *Radiat Med* 2004; **22**: 275-282 [PMID: 15468951]

4 **Takayama Y**, Nishie A, Asayama Y, Ishigami K, Kakihara D, Ushijima Y, Fujita N, Shirabe K, Takemura A, Honda H. Image quality and diagnostic performance of free-breathing diffusion-weighted imaging for hepatocellular carcinoma. *World J Hepatol* 2017; **9**: 657-666 [PMID: 28588750 DOI: 10.4254/wjh.v9.i14.657]

5 **Park MJ**, Kim YK, Choi SY, Rhim H, Lee WJ, Choi D. Preoperative detection of small pancreatic carcinoma: value of adding diffusion-weighted imaging to conventional MR imaging for improving confidence level. *Radiology* 2014; **273**: 433-443 [PMID: 24991989 DOI: 10.1148/radiol.14132563]

6 **Takakura K**, Sumiyama K, Munakata K, Ashida H, Arihiro S, Kakutani H, Tajiri H. Clinical usefulness of diffusion-weighted MR imaging for detection of pancreatic cancer: comparison with enhanced multidetector-row CT. *Abdom Imaging* 2011; **36**: 457-462 [PMID: 21643939 DOI: 10.1007/s00261-011-9728-7]

7 **Jiang D**, Dou W, Vosters L, Xu X, Sun Y, Tan T. Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. *Jpn J Radiol* 2018; **36**: 566-574 [PMID: 29982919 DOI: 10.1007/s11604-018-0758-8]

8 **Kim H**, Lee JM, Yoon JH, Jang JY, Kim SW, Ryu JK, Kannengiesser S, Han JK, Choi BI. Reduced Field-of-View Diffusion-Weighted Magnetic Resonance Imaging of the Pancreas: Comparison with Conventional Single-Shot Echo-Planar Imaging. *Korean J Radiol* 2015; **16**: 1216-1225 [PMID: 26576110 DOI: 10.3348/kjr.2015.16.6.1216]

9 **Ogawa R**, Kido T, Nakamura M, Nozaki A, Lebel RM, Mochizuki T. Reconstruction of cardiovascular black-blood T2-weighted image by deep learning algorithm: A comparison with intensity filter. *Acta Radiol Open* 2021; **10**: 20584601211044779 [PMID: 34594576 DOI: 10.1177/20584601211044779]

10 **Tanabe M**, Higashi M, Benkert T, Imai H, Miyoshi K, Kameda F, Ariyoshi S, Ihara K, Ito K. Reduced Field-of-View Diffusion-Weighted Magnetic Resonance Imaging of the Pancreas With Tilted Excitation Plane: A Preliminary Study. *J Magn Reson Imaging* 2021; **54**: 715-720 [PMID: 33704860 DOI: 10.1002/jmri.27590]

11 **Saritas EU**, Cunningham CH, Lee JH, Han ET, Nishimura DG. DWI of the spinal cord with reduced FOV single-shot EPI. *Magn Reson Med* 2008; **60**: 468-473 [PMID: 18666126 DOI: 10.1002/mrm.21640]

12 **Mannelli L**, Monti S, Corrias G, Fung MM, Nyman C, Golia Pernicka JS, Do RKG. Comparison of Navigator Triggering Reduced Field of View and Large Field of View Diffusion-Weighted Imaging of the Pancreas. *J Comput Assist Tomogr* 2019; **43**: 143-148 [PMID: 30119065 DOI: 10.1097/RCT.0000000000000778]

13 **Warndahl BA**, Borisch EA, Kawashima A, Riederer SJ, Froemming AT. Conventional vs. reduced field of view diffusion weighted imaging of the prostate: Comparison of image quality, correlation with histology, and inter-reader agreement. *Magn Reson Imaging* 2018; **47**: 67-76 [PMID: 29103977 DOI: 10.1016/j.mri.2017.10.011]

14 **Donati F**, Casini C, Cervelli R, Morganti R, Boraschi P. Diffusion-weighted MRI of solid pancreatic lesions: Comparison between reduced field-of-view and large field-of-view sequences. *Eur J Radiol* 2021; **143**: 109936 [PMID: 34464906 DOI: 10.1016/j.ejrad.2021.109936]

15 **Yasaka K**, Akai H, Kunimatsu A, Kiryu S, Abe O. Deep learning with convolutional neural network in radiology. *Jpn J Radiol* 2018; **36**: 257-272 [PMID: 29498017 DOI: 10.1007/s11604-018-0726-3]

16 **Wang X**, Ma J, Bhosale P, Ibarra Rovira JJ, Qayyum A, Sun J, Bayram E, Szklaruk J. Novel deep learning-based noise reduction technique for prostate magnetic resonance imaging. *Abdom Radiol (NY)* 2021; **46**: 3378-3386 [PMID: 33580348 DOI: 10.1007/s00261-021-02964-6]

17 **Baştanlar Y**, Ozuysal M. Introduction to machine learning. *Methods Mol Biol* 2014; **1107**: 105-128 [PMID: 24272434 DOI: 10.1007/978-1-62703-748-8_7]

18 **Nilsson TH**, Nelson TM. Delayed monochromatic hue matches indicate characteristics of visual memory. *J Exp Psychol Hum Percept Perform* 1981; **7**: 141-150 [PMID: 6452491 DOI: 10.1037//0096-1523.7.1.141]

19 **Kim M**, Kim HS, Kim HJ, Park JE, Park SY, Kim YH, Kim SJ, Lee J, Lebel MR. Thin-Slice Pituitary MRI with Deep Learning-based Reconstruction: Diagnostic Performance in a Postoperative Setting. *Radiology* 2021; **298**: 114-122 [PMID: 33141001 DOI: 10.1148/radiol.2020200723]

20 **Kidoh M**, Shinoda K, Kitajima M, Isogawa K, Nambu M, Uetani H, Morita K, Nakaura T, Tateishi M, Yamashita Y. Deep Learning Based Noise Reduction for Brain MR Imaging: Tests on Phantoms and Healthy Volunteers. *Magn Reson Med Sci* 2020; **19**: 195-206 [PMID: 31484849 DOI: 10.2463/mrms.mp.2019-0018]

21 **Bouhrara M**, Maring MC, Spencer RG. A simple and fast adaptive nonlocal multispectral filtering algorithm for efficient noise reduction in magnetic resonance imaging. *Magn Reson Imaging* 2019; **55**: 133-139 [PMID: 30149058 DOI: 10.1016/j.mri.2018.08.011]

22 **Zhu J**, Zhang J, Gao JY, Li JN, Yang DW, Chen M, Zhou C, Yang ZH. Apparent diffusion coefficient normalization of normal liver: Will it improve the reproducibility of diffusion-weighted imaging at different MR scanners as a new biomarker? *Medicine (Baltimore)* 2017; **96**: e5910 [PMID: 28099354 DOI: 10.1097/MD.0000000000005910]

23 **Jones DK**, Basser PJ. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. *Magn Reson Med* 2004; **52**: 979-993 [PMID: 15508154 DOI: 10.1002/mrm.20283]

World Journal of Radiology

Submit a Manuscript: https://www.f6publishing.com

World J Radiol 2023 December 28; 15(12): 350-358

DOI: 10.4329/wjr.v15.i12.350

ISSN 1949-8470 (online)

ORIGINAL ARTICLE

Observational Study

Factors associated with gastrointestinal stromal tumor rupture and pathological risk: A single-center retrospective study

Jia-Zheng Liu, Zhong-Wen Jia, Ling-Ling Sun

Specialty type: Radiology, nuclear medicine and medical imaging

Provenance and peer review: Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): D
Grade E (Poor): E

P-Reviewer: Dimofte GM, Romania; Nishida T, Japan

Received: August 26, 2023
Peer-review started: August 26, 2023
First decision: October 9, 2023
Revised: October 26, 2023
Accepted: December 12, 2023
Article in press: December 12, 2023
Published online: December 28, 2023

Jia-Zheng Liu, Zhong-Wen Jia, Ling-Ling Sun, Department of Radiology, The Fourth Affiliated Hospital of China Medical University, Shenyang 110033, Liaoning Province, China

Corresponding author: Zhong-Wen Jia, PhD, Chief Physician, Department of Radiology, The Fourth Affiliated Hospital of China Medical University, No. 4 Chongshan East Road, Huanggu District, Shenyang 110033, Liaoning Province, China. jiazhongwen@126.com

Abstract

BACKGROUND

Gastrointestinal stromal tumor (GIST) is a rare gastrointestinal mesenchymal tumor with potential malignancy. Once the tumor ruptures, regardless of tumor size and mitotic number, it can be identified into a high-risk group. It is of great significance for the diagnosis, treatment, and prognosis of GIST if non-invasive examination can be performed before surgery to accurately assess the risk of tumor.

AIM

To identify the factors associated with GIST rupture and pathological risk.

METHODS

A cohort of 50 patients with GISTs, as confirmed by postoperative pathology, was selected from our hospital. Clinicopathological and computed tomography data of the patients were collected. Logistic regression analysis was used to evaluate factors associated with GIST rupture and pathological risk grade.

RESULTS

Pathological risk grade, tumor diameter, tumor morphology, internal necrosis, gas-liquid interface, and Ki-67 index exhibited significant associations with GIST rupture ($P < 0.05$). Gender, tumor diameter, tumor rupture, and Ki-67 index were found to be correlated with pathological risk grade of GIST ($P < 0.05$). Multifactorial logistic regression analysis revealed that male gender and tumor diameter ≥ 10 cm were independent predictors of a high pathological risk grade of GIST [odds ratio (OR) = 11.12, 95% confidence interval (95%CI): 1.81-68.52, $P = 0.01$; OR = 22.96, 95%CI: 2.19-240.93, $P = 0.01$]. Tumor diameter ≥ 10 cm, irregular shape, internal necrosis, gas-liquid interface, and Ki-67 index ≥ 10 were identified as independent predictors of a high risk of GIST rupture (OR = 9.67, 95%CI: 2.15-43.56, $P = 0.01$; OR = 35.44, 95%CI: 4.01-313.38, $P < 0.01$; OR = 18.75, 95%CI: 3.40-103.34, $P < 0.01$; OR = 27.00, 95%CI: 3.10-235.02, $P < 0.01$; OR = 4.43, 95%CI: 1.10-

17.92, $P = 0.04$).

CONCLUSION

Tumor diameter, tumor morphology, internal necrosis, gas-liquid, and Ki-67 index are associated with GIST rupture, while gender and tumor diameter are linked to the pathological risk of GIST. These findings contribute to our understanding of GIST and may inform non-invasive examination strategies and risk assessment for this condition.

Key Words: Gastrointestinal stromal tumors; Imaging findings; Tumor rupture; Pathological risk grades

Core Tip: Gastrointestinal stromal tumor (GIST) biopsy is inconvenient, has a low yield, and easily leads to tumor metastasis. It is of great significance for the diagnosis, treatment, and prognosis of GIST if non-invasive examination can be performed before surgery to accurately assess the risk of tumor. The results of our study found that tumor diameter, tumor morphology, internal necrosis, and gas-liquid interface are related to the rupture of GIST, and sex and tumor diameter are related to the pathological risk of GIST. The results of this study provides ideas for non-invasive examination and risk assessment of GIST.

Citation: Liu JZ, Jia ZW, Sun LL. Factors associated with gastrointestinal stromal tumor rupture and pathological risk: A single-center retrospective study. *World J Radiol* 2023; 15(12): 350-358
URL: https://www.wjgnet.com/1949-8470/full/v15/i12/350.htm
DOI: https://dx.doi.org/10.4329/wjr.v15.i12.350

INTRODUCTION

Gastrointestinal stromal tumor (GIST), a rare mesenchymal tumor of the gastrointestinal tract, presents a potential for malignancy and constitutes 1%-3% of gastrointestinal malignancies[1,2]. Immunohistochemical analysis of GIST typically reveals positive expression of CD117, CD34, or DOG-1[3,4]. Due to its invasive nature and propensity for recurrence and metastasis, the clinical assessment of prognosis following GIST surgery heavily relies on pathological evaluation. However, preoperative selection of appropriate treatment methods lacks a foundation based on pathological assessment. Notably, imaging characteristics of GIST have been observed, and significant disparities in postoperative pathological risk grades have been identified between GISTs exhibiting distinct computed tomography (CT) features prior to surgery, thereby highlighting the crucial role of CT in GIST diagnosis[5,6].

GISTs display unpredictable and variable biological behavior, rendering the distinction between benign and malignant tumors challenging[2,7]. In the early stages, GISTs were classified as either benign or malignant; however, clinical experience has revealed that tumors initially determined as "benign" by histopathology may later metastasize. Consequently, many pathologists advocate for grouping based on pathological risk grades[8,9]. Once the tumor ruptures, irrespective of size and mitotic count, it can be classified into a high-risk group.

GIST biopsy is inconvenient and has a limited yield, and open biopsies can potentially induce tumor metastasis, precluding risk assessment in such cases. Risk assessment cannot be performed for biopsied cases. Therefore, needle biopsy is not recommended prior to surgery for GISTs that can be completely resected[10]. Given the divergent treatment and prognosis of GISTs compared to non-epithelial tumors like lymphoma and schwannoma, preoperative imaging diagnosis and evaluation assume paramount importance. The ability to perform non-invasive examinations before surgery to accurately assess tumor risk would hold significant implications for GIST diagnosis, treatment, and prognosis. In light of this, we postulated that imaging findings possess clinical utility in predicting GIST rupture and pathological risk. Consequently, this study aimed to offer insights into non-invasive examination strategies and risk assessment for GISTs by examining the correlation between imaging findings and GIST rupture and pathological risk.

MATERIALS AND METHODS

Patients

Fifty patients diagnosed with GISTs were included in this retrospective study, following confirmation of the diagnosis through postoperative pathology at our institution. The patients' clinicopathological and CT data were systematically collected. The study cohort consisted of individuals aged between 18 and 84 years, comprising 28 males and 22 females. In order to ensure the reliability and relevance of the data, specific inclusion and exclusion criteria were applied. The inclusion criteria encompassed patients who had undergone biopsy or surgery at our hospital, with complete and well-documented pathological data, clear risk grading, and comprehensive clinical and CT data available. Furthermore, only

primary tumors were considered. Patients who had not undergone CT examination prior to surgery or whose CT image quality was deemed inadequate were excluded. Additionally, cases with uncertain tumor pathological risk grading or those involving tumor relapse were also excluded from the study cohort.

Data collected

In this investigation, we meticulously gathered a comprehensive set of clinical and pathological data from a cohort of 50 patients diagnosed with GISTs. The dataset encompassed crucial patient demographics such as age and gender, as well as pivotal pathological indicators including risk grade, tumor diameter, morphology, necrosis, rupture status, gas-liquid interface, tumor location, mitotic figures, and Ki-67 index. The assessment of pathological risk was meticulously categorized into four distinct levels, namely, very low, low, moderate, and high, enabling a comprehensive evaluation of the disease severity[8,11,12] (Supplementary Table 1).

CT scanning and indicators

Contrast-enhanced CT scanning was performed using a 256-slice computed tomography scanner (Brilliance iCT, Philips) with the following scanning conditions: Peak kilovoltage of 120 and tube current (mA) ranging from 138 to 458. The following parameters were assessed: (1) Tumor diameter: The maximum diameter of the tumor was measured on the coronal image; (2) Tumor morphology: The shape of the tumor was evaluated to determine if it exhibited a regular shape. A tumor with an elliptical or round shape was considered regular; (3) Boundary: The boundary of the tumor was assessed based on the presence of a clear boundary or an unclear boundary. An unclear boundary indicated a potential for invasion; (4) Primary tumor site: The primary tumor site was determined based on the location of the initial lesion; (5) Necrosis: The presence of a necrotic area was determined based on the CT results; and (6) Gas-liquid interface: The presence of a gas-liquid interface was assessed based on the imaging results. These parameters were evaluated to assess the risk factors associated with GIST rupture and pathological risk.

Criteria for tumor rupture

The criteria for tumor rupture included: (1) Tumor rupture or overflow; (2) Presence of bloody ascites; (3) Gastrointestinal perforation at the tumor site; (4) Microscopic infiltration of adjacent organs; (5) Intra-lesional dissection or segmental resection; and (6) Incisional biopsy[12,13].

Statistical analysis

SPSS 26.0 (IBM Corp, Armonk, NY) software was used for statistical analyses. Enumeration data are expressed as frequencies, and statistical analysis was performed by the χ^2 test. Pearson correlation was used to analyze the correlation between age, gender, pathological risk grade, tumor diameter, tumor morphology, internal necrosis, tumor rupture, gas-liquid interface, tumor site, mitotic figures, and Ki-67 index. $P < 0.05$ was considered statistically significant.

RESULTS

Analysis of related factors of GIST rupture

The results of the comparison of clinical data between the unruptured and ruptured GISTs are shown in Table 1. Statistical analysis showed that pathological risk, tumor diameter, tumor morphology, internal necrosis, and gas-liquid interface were associated with GIST rupture ($P < 0.05$). The differences in age, gender, primary site, mitotic count, and Ki-67 index of the ruptured group and the unruptured group were not statistically significant ($P > 0.05$). GISTs with a high pathological risk grade, large tumor diameter, irregular shape, internal tumor necrosis, and gas-fluid interface were prone to rupture.

Analysis of risk factors in GIST patients

The pathological risk grade assessment of GISTs was carried out through various observation indicators of CT images. The results showed that there were 24 cases of low risk, 6 cases of intermediate risk, and 20 cases of high risk. The analysis results showed that gender, tumor diameter, tumor rupture, and Ki-67 index were associated with the pathological risk grade of GISTs ($P < 0.05$) (Table 2). We found that male GIST patients had a higher pathological risk grade, and the longer the tumor diameter, the higher the pathological risk of GISTs. GIST patients whose tumors were prone to rupture had a higher pathological risk grade, multiple gas shadows were common in the central necrotic area of ruptured tumors (Figure 1A), gas-liquid interface (Figure 1B) was visible in the tumor, and pus coating was formed next to the tumor (Figure 1C).

Logistic regression analysis of factors associated with pathologic risk grade of GISTs

The results of the logistic regression analysis of the factors associated with the pathological risk grade of GISTs are shown in Table 3. Multifactorial logistic regression analysis showed that male gender and tumor diameter ≥ 10 cm were independently correlated with a high pathological risk grade of GISTs [odds ratio (OR) = 11.12, 95% confidence interval (95%CI): 1.81-68.52, $P = 0.01$; OR = 22.96, 95%CI: 2.19-240.93, $P = 0.01$].

Table 1 Comparison of clinical data between unruptured and ruptured gastrointestinal stromal tumors

	Unruptured group (*n* = 38)	Rupture group (*n* = 12)	Statistical value	*P* value
Age (yr)	64.79 ± 9.75	57.17 ± 20.61	1.24	0.24
Gender			2.31	0.13
Male	19	9		
Female	19	3		
Pathological risk grade			8.47	0.02
Low risk	21	3		
Intermediate risk	6	0		
High risk	11	9		
Tumor diameter (cm)	3.25 (2, 5.25)	9.0 (4.5, 13.63)	2.60	0.01
Tumor shape			17.56	< 0.01
Irregular	9	11		
Regular	29	1		
Internal necrosis			15.35	< 0.01
No	30	2		
Yes	8	10		
Gas-liquid interface			23.68	< 0.01
Absence	27	1		
Presence	11	11		
Primary site			0.23	0.63
Gastric	22	6		
Small bowel	16	6		
Mitotic count (50/HPF)	6.11 ± 2.60	6.17 ± 1.64	0.08	0.94
Ki-67 index (%)	5.50 (4.75, 8.00)	7.50 (5.00, 14.25)	1.15	0.25

HPF: High-power field.

Logistic regression analysis of factors associated with GIST rupture

The results of the logistic regression analysis of the factors associated with the tumor rupture of GIST are shown in Table 4. Multifactorial logistic regression analysis showed that tumor diameter ≥ 10 cm, irregular shape, internal necrosis, gas-liquid interface, and Ki-67 index ≥ 10 were independently correlated with a high risk of tumor rupture of GISTs (OR = 9.67, 95%CI: 2.15-43.56, $P = 0.01$; OR = 35.44, 95%CI: 4.01-313.38, $P < 0.01$; OR = 18.75, 95%CI: 3.40-103.34, $P < 0.01$; OR = 27.00, 95%CI: 3.10-235.02, $P < 0.01$; OR = 4.43, 95%CI: 1.10-17.92, $P = 0.04$).

DISCUSSION

In this study, our findings indicated that certain factors are associated with the rupture of GISTs in the patients that we screened. These factors include tumor diameter, tumor shape, internal necrosis, and gas-liquid interface. Additionally, we found that being male and having a tumor diameter ≥ 10 cm are independent correlates of a high pathological risk grade of GISTs.

GIST is a gastrointestinal tumor that has seen a significant increase in the incidence and diagnosis rate in recent years. Rupture and bleeding of GISTs are considered to be serious and dangerous complications that require urgent attention[2,14]. The clinical manifestations of spontaneous tumor rupture and hemorrhage are atypical, characterized by a rapid onset. Many patients are admitted to the hospital with acute abdomen, resulting in delayed surgery. Therefore, timely diagnosis and treatment are crucial for improving patient prognosis[15,16].

Tumor rupture is an important risk factor for recurrence after GIST resection and is also an indicator for adjuvant imatinib therapy[13]. Numerous studies have confirmed that tumor rupture is associated with an increased risk of recurrence. For example, Yanagimoto *et al*[17] identified that tumor size, mitotic count, tumor location, and tumor rupture were important prognostic factors for GIST. Hølmebakk *et al*[18] and Nishida *et al*[19] found that tumor rupture was an

Table 2 Analysis of related factors of pathological risk grade of gastrointestinal stromal tumors

	Low risk (*n* = 24)	Intermediate risk (*n* = 6)	High risk (*n* = 20)	χ^2 value	*P* value
Age (yr)				1.94	0.38
< 60	5	2	8		
≥ 60	19	4	12		
Gender				8.10	0.02
Male	9	3	16		
Female	15	3	4		
Tumor diameter (cm)				10.47	0.01
< 10	22	6	11		
≥ 10	2	0	9		
Tumor shape				2.26	0.32
Irregular	9	1	10		
Regular	15	5	10		
Internal necrosis				5.37	0.07
No	18	5	9		
Yes	6	1	11		
Tumor rupture				8.47	0.02
No	21	6	11		
Yes	3	0	9		
Gas-liquid interface				0.79	0.67
Absence	15	3	10		
Presence	9	3	10		
Primary site				3.46	0.18
Gastric	16	4	8		
Small bowel	8	2	12		
Mitotic count (50/HPF)				4.49	0.11
< 5	8	3	3		
≥ 5	16	2	17		
Ki-67 index (%)				6.30	0.04
< 10	21	5	11		
≥ 10	3	1	9		

HPF: High-power field.

independent prognostic factor for recurrence-free survival. These findings highlight the significance of tumor rupture in evaluating the prognosis of GIST patients and its association with the poor outcomes.

Furthermore, approximately half of GIST ruptures are spontaneous and cannot be prevented. Therefore, there is growing interest in studying factors related to tumor rupture[19-22]. Our study identified tumor diameter, tumor shape, internal necrosis, and the presence of gas-liquid interface as factors associated with GIST rupture. Previous research has also reported that larger tumor diameters are associated with a higher risk of rupture[19], and that larger tumors are more likely to experience necrosis in the central region[23]. Positive resection margins have also been strongly linked to tumor rupture[24]. Moreover, the clinical presentation of GISTs, such as an unclear tumor boundary, irregular tumor shape, and the presence of a gas-liquid interface in imaging scans, can indicate aggressive behavior and malignancy. Gas-liquid interface detection in GISTs is currently uncommon, but researchers believe that it predicts severe disease in GIST patients[25-27]. Our study found that necrosis and rupture were more likely to occur when an air-liquid interface was present, and these factors were important indicators of poor prognosis in GIST patients. However, it is important to note that the definition of tumor rupture remains controversial, and consistent standards have yet to be established[18]. Some

Table 3 Logistic regression analysis of factors associated with the pathological risk grade of gastrointestinal stromal tumors

	B	S.E.	Wald	Sig.	Exp (B)	95%CI for Exp (B)	
						Lower	Upper
Age ≥ 60 yr	0.44	0.90	0.24	0.63	1.55	0.27	9.07
Male gender	2.41	0.93	6.75	0.01	11.12	1.81	68.52
Tumor diameter ≥ 10 cm	3.13	1.20	6.83	0.01	22.96	2.19	240.93
Irregular shape	0.50	1.12	0.20	0.65	1.65	0.18	14.92
Internal necrosis	-1.40	1.21	1.35	0.25	0.25	0.02	2.62
Gas-liquid interface	0.50	1.26	0.16	0.69	1.64	0.14	19.48
Small bowel tumor	-0.15	0.87	0.03	0.87	0.87	0.16	4.76
50/HPF ≥ 5	-0.33	0.87	0.14	0.71	0.72	0.13	3.96
Ki-67 index ≥ 10	2.18	1.15	3.56	0.06	8.82	0.92	84.49

95%CI: 95% confidence interval; HPF: High-power field.

Table 4 Logistic regression analysis of factors associated with rupture of gastrointestinal stromal tumors

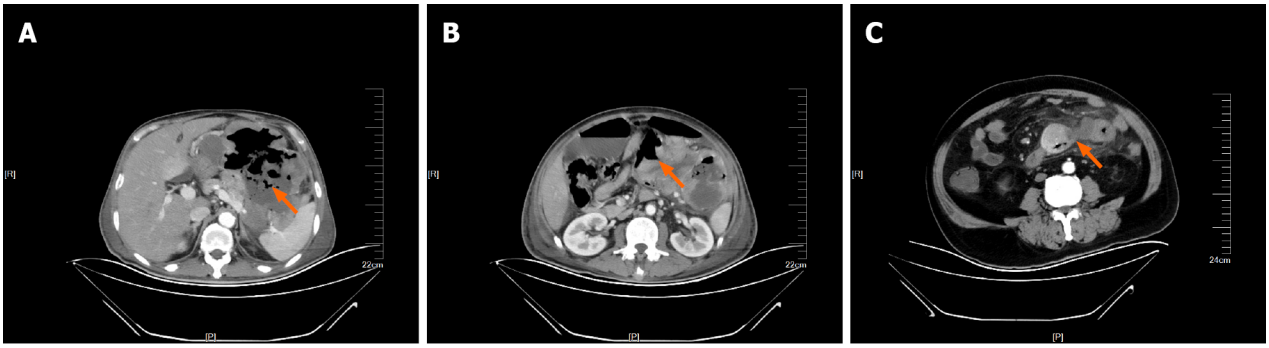
	B	S.E.	Wald	Sig.	Exp (B)	95%CI for Exp (B)	
						Lower	Upper
Age ≥ 60 yr	0.21	0.71	0.08	0.77	1.23	0.31	4.93
Male gender	-1.10	0.74	2.20	0.14	0.33	0.08	1.43
Tumor diameter ≥ 10 cm	2.27	0.77	8.72	< 0.01	9.67	2.15	43.56
Irregular shape	3.57	1.11	10.30	< 0.01	35.44	4.01	313.38
Internal necrosis	2.93	0.87	11.33	< 0.01	18.75	3.40	103.34
Gas-liquid interface	3.30	1.10	8.91	< 0.01	27.00	3.10	235.02
Small bowel tumor	-0.32	0.66	0.23	0.63	0.73	0.20	2.67
50/HPF ≥ 5	-0.84	0.85	0.97	0.33	0.43	0.08	2.29
Ki-67 index ≥ 10	1.49	0.71	4.36	0.04	4.43	1.10	17.92
High pathological risk grade	-1.31	0.74	3.12	0.08	0.27	0.06	1.16

95%CI: 95% confidence interval; HPF: High-power field.

researchers consider macroscopic damage of tumor pseudocapsule as tumor rupture[19].

In cases of GISTs with high pathological risk grades, CT signs of malignancy include invasive tumor growth, large size with uneven density and unclear boundaries, hemorrhage, liquefaction, necrosis or cystic degeneration, inhomogeneous enhancement on CT enhancement, and the presence of thick tumor blood vessels around the tumor in the arterial phase. Additionally, GISTs metastasizing to other organs and extra-GISTs located outside the gastrointestinal tract are prone to malignancy. Our study found that gender, tumor diameter, rupture, and Ki-67 index were closely associated with pathological risk grades. Lower pathological risk grades of GIST are characterized by slow tumor growth, smaller tumor diameters (usually less than 5.0 cm), round or oval shapes, uniform enhancement on scans, no invasion of surrounding tissues, and no distant organ metastasis. Conversely, higher pathological risk grades indicate worse growth and larger tumor diameters. These tumors are more likely to experience liquefaction and necrosis due to a relative lack of blood supply. Our study suggests that combining CT examination with tumor diameter, morphology, internal necrosis, gas-liquid interface, and Ki-67 index can facilitate early non-invasive assessment of GIST tumor rupture risk, providing valuable information for clinical decision-making. Additionally, clinical diagnostic information can be used to predict the pathological risk grades of GISTs, aiding in further clinical diagnosis and treatment.

There are some limitations to this study that should be acknowledged. First, the small number of GIST samples included warrants further studies with larger sample sizes. Second, the study primarily focused on GIST cases occurring in the gastric and small bowel, which may not fully reflect the relationship between tumor location and tumor rupture and pathological risk grade. Therefore, it is necessary to include more GIST cases in uncommon sites. Lastly, the study lacks information on treatment modalities and the presence of metastases.



DOI: 10.4329/wjr.v15.i12.350

Figure 1 Computed tomography images of ruptured gastrointestinal stromal tumors. A: Rupture of a giant gastrointestinal stromal tumor at the colosplenic flexure, with multiple gas shadows (arrow) seen in the central necrotic zone; B: Rupture of a small intestinal stromal tumor. The arrow points to the ruptured opening of the tumor, and the gas-liquid interface was seen in the abdominal cavity; C: A ruptured small intestinal stromal tumor. Gas can be seen in the tumor, and the arrow points to the formation of pus coating around the tumor.

CONCLUSION

In summary, our study has substantiated the association between tumor diameter, tumor shape, internal necrosis, and gas-liquid interface with the occurrence of GIST rupture. Furthermore, we have identified gender and tumor diameter as independent factors influencing the pathological risk grade of GISTs. By leveraging the power of CT detection and integrating the aforementioned factors, we have successfully demonstrated the potential of non-invasive early assessment for GIST rupture and pathological risk grade. These findings hold significant promise in enhancing the clinical decision-making process by providing valuable insights.

ARTICLE HIGHLIGHTS

Research background

Gastrointestinal stromal tumor (GIST) is a rare gastrointestinal mesenchymal tumor. It is of great significance for the diagnosis, treatment, and prognosis of GIST if non-invasive examination can be performed before surgery to accurately assess the risk of tumor.

Research motivation

If accurate assessment of GIST tumor risk through non-invasive examination is the focus of this study, it can provide valuable insights into non-invasive examination strategies and risk assessment of GISTs.

Research objectives

To investigate the factors associated with GIST rupture and pathological risk, and provide insights into non-invasive examination techniques and risk assessment for GISTs.

Research methods

A cohort of 50 GIST patients was selected from our hospital. Clinicopathological and CT data of the patients were collected. Logistic regression analysis was used to evaluate factors associated with GIST rupture and pathological risk grade.

Research results

Male gender and tumor diameter ≥ 10 cm were independent predictors of a high pathological risk grade of GISTs [odds ratio (OR) = 11.12, 95% confidence interval (95%CI): 1.81-68.52, P = 0.01; OR = 22.96, 95%CI: 2.19-240.93, P = 0.01]. Tumor diameter ≥ 10 cm, irregular shape, internal necrosis, gas-liquid interface, and Ki-67 index ≥ 10 were identified as independent predictors of a high risk of GIST rupture (OR = 9.67, 95%CI: 2.15-43.56, P = 0.01; OR = 35.44, 95%CI: 4.01-313.38, P < 0.01; OR = 18.75, 95%CI: 3.40-103.34, P < 0.01; OR = 27.00, 95%CI: 3.10-235.02, P < 0.01; OR = 4.43, 95%CI: 1.10-17.92, P = 0.04).

Research conclusions

Tumor diameter, tumor morphology, internal necrosis, gas-liquid interface, and Ki-67 index are associated with GIST rupture, while gender and tumor diameter are linked to the pathological risk of GISTs. These findings contribute to our understanding of GISTs and may inform non-invasive examination strategies and risk assessment for this condition.

Research perspectives

In later studies, we can further verify our conclusions in large-sample clinical studies to better guide clinical non-invasive examination and risk assessment of GISTs.

FOOTNOTES

Author contributions: Jia ZW conceived the idea; Liu JZ and Sun LL collected and analyzed the data; Liu JZ wrote the paper; all the authors discussed the results and contributed to the final manuscript.

Institutional review board statement: The study was reviewed and approved by The Fourth Affiliated Hospital of China Medical University Institutional Review Board.

Informed consent statement: Patients were not required to give informed consent to the study because the analysis used anonymous clinical data that were obtained after each patient agreed to treatment by written consent.

Conflict-of-interest statement: All the authors have no conflicts of interest to disclose.

Data sharing statement: The data used to support the findings of this study are available from the corresponding author upon request.

STROBE statement: The authors have read the STROBE Statement—checklist of items, and the manuscript was prepared and revised according to the STROBE Statement—checklist of items.

Country/Territory of origin: China

ORCID number: Zhong-Wen Jia 0009-0003-3835-1666.

S-Editor: Lin C
L-Editor: Wang TQ
P-Editor: Lin C

REFERENCES

1 Gastrointestinal stromal tumours. *Nat Rev Dis Primers* 2021; **7**: 21 [PMID: 33737509 DOI: 10.1038/s41572-021-00262-5]
2 **Ho MY**, Blanke CD. Gastrointestinal stromal tumors: disease and treatment update. *Gastroenterology* 2011; **140**: 1372-6.e2 [PMID: 21420965 DOI: 10.1053/j.gastro.2011.03.017]
3 **Koo DH**, Ryu MH, Kim KM, Yang HK, Sawaki A, Hirota S, Zheng J, Zhang B, Tzen CY, Yeh CN, Nishida T, Shen L, Chen LT, Kang YK. Asian Consensus Guidelines for the Diagnosis and Management of Gastrointestinal Stromal Tumor. *Cancer Res Treat* 2016; **48**: 1155-1166 [PMID: 27384163 DOI: 10.4143/crt.2016.187]
4 **Qi ZP**, Shi Q, Liu JZ, Yao LQ, Xu MD, Cai SL, Li B, Take I, Zhang YQ, Chen WF, Zhong YS, Zhou PH. Efficacy and safety of endoscopic submucosal dissection for submucosal tumors of the colon and rectum. *Gastrointest Endosc* 2018; **87**: 540-548.e1 [PMID: 28987548 DOI: 10.1016/j.gie.2017.09.027]
5 **Choi YR**, Kim SH, Kim SA, Shin CI, Kim HJ, Han JK, Choi BI. Differentiation of large (≥ 5 cm) gastrointestinal stromal tumors from benign subepithelial tumors in the stomach: radiologists' performance using CT. *Eur J Radiol* 2014; **83**: 250-260 [PMID: 24325848 DOI: 10.1016/j.ejrad.2013.10.028]
6 **Teoh WC**, Teo SY, Ong CL. Gastrointestinal stromal tumors presenting as gynecological masses: usefulness of multidetector computed tomography. *Ultrasound Obstet Gynecol* 2011; **37**: 107-109 [PMID: 20737452 DOI: 10.1002/uog.8801]
7 **Ibrahim A**, Chopra S. Succinate Dehydrogenase-Deficient Gastrointestinal Stromal Tumors. *Arch Pathol Lab Med* 2020; **144**: 655-660 [PMID: 31169996 DOI: 10.5858/arpa.2018-0370-RS]
8 **Raut CP**, Espat NJ, Maki RG, Araujo DM, Trent J, Williams TF, Purkayastha DD, DeMatteo RP. Efficacy and Tolerability of 5-Year Adjuvant Imatinib Treatment for Patients With Resected Intermediate- or High-Risk Primary Gastrointestinal Stromal Tumor: The PERSIST-5 Clinical Trial. *JAMA Oncol* 2018; **4**: e184060 [PMID: 30383140 DOI: 10.1001/jamaoncol.2018.4060]
9 **Chu H**, Pang P, He J, Zhang D, Zhang M, Qiu Y, Li X, Lei P, Fan B, Xu R. Value of radiomics model based on enhanced computed tomography in risk grade prediction of gastrointestinal stromal tumors. *Sci Rep* 2021; **11**: 12009 [PMID: 34103619 DOI: 10.1038/s41598-021-91508-5]
10 **Wei ZG**, Han C, Li JD. Treatment Experts Group for Gastrointestinal Stromal. *Shijie Huaren Xiaohua Zazhi* 2010; **18**: 65-69 [DOI: 10.11569/wcjd.v18.i1.65]
11 **Wei SC**, Xu L, Li WH, Li Y, Guo SF, Sun XR, Li WW. Risk stratification in GIST: shape quantification with CT is a predictive factor. *Eur Radiol* 2020; **30**: 1856-1865 [PMID: 31900704 DOI: 10.1007/s00330-019-06561-6]
12 **Fletcher CD**, Berman JJ, Corless C, Gorstein F, Lasota J, Longley BJ, Miettinen M, O'Leary TJ, Remotti H, Rubin BP, Shmookler B, Sobin

LH, Weiss SW. Diagnosis of gastrointestinal stromal tumors: A consensus approach. *Hum Pathol* 2002; **33**: 459-465 [PMID: 12094370 DOI: 10.1053/hupa.2002.123545]

13 **Nishida T**, Hølmebakk T, Raut CP, Rutkowski P. Defining Tumor Rupture in Gastrointestinal Stromal Tumor. *Ann Surg Oncol* 2019; **26**: 1669-1675 [PMID: 30868512 DOI: 10.1245/s10434-019-07297-9]

14 **Faigel DO**, Abulhawa S. Gastrointestinal stromal tumors: the role of the gastroenterologist in diagnosis and risk stratification. *J Clin Gastroenterol* 2012; **46**: 629-636 [PMID: 22858511 DOI: 10.1097/MCG.0b013e3182548f6c]

15 **Blay JY**, Kang YK, Nishida T, von Mehren M. Gastrointestinal stromal tumours. *Nat Rev Dis Primers* 2021; **7**: 22 [PMID: 33737510 DOI: 10.1038/s41572-021-00254-5]

16 **Fan X**, Han H, Sun Z, Zhang L, Chen G, Mzee SAS, Yang H, Chen J. Prognostic Value of Bleeding in Gastrointestinal Stromal Tumors: A Meta-Analysis. *Technol Cancer Res Treat* 2021; **20**: 15330338211034259 [PMID: 34323156 DOI: 10.1177/15330338211034259]

17 **Yanagimoto Y**, Takahashi T, Muguruma K, Toyokawa T, Kusanagi H, Omori T, Masuzawa T, Tanaka K, Hirota S, Nishida T. Re-appraisal of risk classifications for primary gastrointestinal stromal tumors (GISTs) after complete resection: indications for adjuvant therapy. *Gastric Cancer* 2015; **18**: 426-433 [PMID: 24853473 DOI: 10.1007/s10120-014-0386-7]

18 **Hølmebakk T**, Hompland I, Bjerkehagen B, Stoldt S, Bruland ØS, Hall KS, Boye K. Recurrence-Free Survival After Resection of Gastric Gastrointestinal Stromal Tumors Classified According to a Strict Definition of Tumor Rupture: A Population-Based Study. *Ann Surg Oncol* 2018; **25**: 1133-1139 [PMID: 29435684 DOI: 10.1245/s10434-018-6353-5]

19 **Nishida T**, Cho H, Hirota S, Masuzawa T, Chiguchi G, Tsujinaka T; Kinki GIST Study Group. Clinicopathological Features and Prognosis of Primary GISTs with Tumor Rupture in the Real World. *Ann Surg Oncol* 2018; **25**: 1961-1969 [PMID: 29752602 DOI: 10.1245/s10434-018-6505-7]

20 **Hacking S**, Jackson K, Johnston R, Grogan E, Walmsley R, Armstrong L, Coleman J, Grieve J, Robinson S, Frew K, Aujayeb A. GIST-related malignant ascites with large-volume paracentesis complicated by myocardial infarction and tumour rupture. *BMJ Support Palliat Care* 2023; **13**: e93-e95 [PMID: 32792419 DOI: 10.1136/bmjspcare-2020-002506]

21 **Misawa S**, Takeda M, Sakamoto H, Kirii Y, Ota H, Takagi H. Spontaneous rupture of a giant gastrointestinal stromal tumor of the jejunum: a case report and literature review. *World J Surg Oncol* 2014; **12**: 153 [PMID: 24885725 DOI: 10.1186/1477-7819-12-153]

22 **Santoyo-Villalba J**, Triguero-Cabrera J, García-Jiménez A. Acute abdomen secondary to massive bleeding due to rupture of liver metastases of a gastrointestinal stroma tumor. *Cir Cir* 2021; **89**: 93-96 [PMID: 34762635 DOI: 10.24875/CIRU.20001313]

23 **Miettinen M**, Lasota J. Gastrointestinal stromal tumors: pathology and prognosis at different sites. *Semin Diagn Pathol* 2006; **23**: 70-83 [PMID: 17193820 DOI: 10.1053/j.semdp.2006.09.001]

24 **Hølmebakk T**, Bjerkehagen B, Hompland I, Stoldt S, Boye K. Relationship between R1 resection, tumour rupture and recurrence in resected gastrointestinal stromal tumour. *Br J Surg* 2019; **106**: 419-426 [PMID: 30507040 DOI: 10.1002/bjs.11027]

25 **Sureka B**, Bansal K, Arora A. Torricelli-Bernoulli Sign in Gastrointestinal Stromal Tumor. *AJR Am J Roentgenol* 2015; **205**: W468 [PMID: 26397359 DOI: 10.2214/AJR.15.14926]

26 **Fortman BJ**. Torricelli-Bernoulli sign in an ulcerating gastric leiomyosarcoma. *AJR Am J Roentgenol* 1999; **173**: 199-200 [PMID: 10397126 DOI: 10.2214/ajr.173.1.10397126]

27 **Liu T**, Lin G, Peng H, Huang L, Jiang X, Li H, Cai K, Jiang J, Guo L, Du X, Tang J, Zhang W, Chen J, Ye Y. Clinicopathological characteristics and prognosis of gastrointestinal stromal tumors containing air-fluid levels. *PLoS One* 2021; **16**: e0261566 [PMID: 34919581 DOI: 10.1371/journal.pone.0261566]

World Journal of Radiology

Submit a Manuscript: https://www.f6publishing.com

World J Radiol 2023 December 28; 15(12): 359-369

DOI: 10.4329/wjr.v15.i12.359

ISSN 1949-8470 (online)

ORIGINAL ARTICLE

Observational Study

Methods for improving colorectal cancer annotation efficiency for artificial intelligence-observer training

Matthew Grudza, Brandon Salinel, Sarah Zeien, Matthew Murphy, Jake Adkins, Corey T Jensen, Curtis Bay, Vikram Kodibagkar, Phillip Koo, Tomislav Dragovich, Michael A Choti, Madappa Kundranda, Tanveer Syeda-Mahmood, Hong-Zhi Wang, John Chang

Specialty type: Radiology, nuclear medicine and medical imaging

Provenance and peer review: Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Fu L, China; Zhu L, China

Received: October 3, 2023
Peer-review started: October 3, 2023
First decision: October 9, 2023
Revised: November 13, 2023
Accepted: December 5, 2023
Article in press: December 5, 2023
Published online: December 28, 2023

Matthew Grudza, School of Biological Health and Systems Engineering, Arizona State University, Tempe, AZ 85287, United States

Brandon Salinel, Phillip Koo, John Chang, Department of Radiology, Banner MD Anderson Cancer Center, Gilbert, AZ 85234, United States

Sarah Zeien, Matthew Murphy, School of Osteopathic Medicine, A.T. Still University, Mesa, AZ 85206, United States

Jake Adkins, Department of Abdominal Imaging, MD Anderson Cancer Center, Houston, TX 77030, United States

Corey T Jensen, Department of Abdominal Imaging, University Texas MD Anderson Cancer Center, Houston, TX 77030, United States

Curtis Bay, Department of Interdisciplinary Sciences, A.T. Still University, Mesa, AZ 85206, United States

Vikram Kodibagkar, School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287, United States

Tomislav Dragovich, Madappa Kundranda, Division of Cancer Medicine, Banner MD Anderson Cancer Center, Gilbert, AZ 85234, United States

Michael A Choti, Department of Surgical Oncology, Banner MD Anderson Cancer Center, Gilbert, AZ 85234, United States

Tanveer Syeda-Mahmood, Hong-Zhi Wang, IBM Almaden Research Center, IBM, San Jose, CA 95120, United States

Corresponding author: John Chang, MD, PhD, Doctor, Department of Radiology, Banner MD Anderson Cancer Center, 2940 E. Banner Gateway Drive, Suite 315, Gilbert, AZ 85234, United States. changresearch1@gmail.com

Abstract

BACKGROUND

Missing occult cancer lesions accounts for the most diagnostic errors in retro-

spective radiology reviews as early cancer can be small or subtle, making the lesions difficult to detect. Second-observer is the most effective technique for reducing these events and can be economically implemented with the advent of artificial intelligence (AI).

AIM

To achieve appropriate AI model training, a large annotated dataset is necessary to train the AI models. Our goal in this research is to compare two methods for decreasing the annotation time to establish ground truth: Skip-slice annotation and AI-initiated annotation.

METHODS

We developed a 2D U-Net as an AI second observer for detecting colorectal cancer (CRC) and an ensemble of 5 differently initiated 2D U-Net for ensemble technique. Each model was trained with 51 cases of annotated CRC computed tomography of the abdomen and pelvis, tested with 7 cases, and validated with 20 cases from The Cancer Imaging Archive cases. The sensitivity, false positives per case, and estimated Dice coefficient were obtained for each method of training. We compared the two methods of annotations and the time reduction associated with the technique. The time differences were tested using Friedman's two-way analysis of variance.

RESULTS

Sparse annotation significantly reduces the time for annotation particularly skipping 2 slices at a time ($P < 0.001$). Reduction of up to 2/3 of the annotation does not reduce AI model sensitivity or false positives per case. Although initializing human annotation with AI reduces the annotation time, the reduction is minimal, even when using an ensemble AI to decrease false positives.

CONCLUSION

Our data support the sparse annotation technique as an efficient technique for reducing the time needed to establish the ground truth.

Key Words: Artificial intelligence; Colorectal cancer; Detection

Core Tip: Minimizing diagnostic errors for colorectal cancer may be most effectively performed with artificial intelligence (AI) second observer. Supervised training of AI-observer will require high volume of annotated training cases. Comparing skip-slice annotation and AI-initiated annotation shows that skipping slices does not affect the training outcome while AI-initiated annotation does not significantly improve annotation time.

Citation: Grudza M, Salinel B, Zeien S, Murphy M, Adkins J, Jensen CT, Bay C, Kodibagkar V, Koo P, Dragovich T, Choti MA, Kundranda M, Syeda-Mahmood T, Wang HZ, Chang J. Methods for improving colorectal cancer annotation efficiency for artificial intelligence-observer training. *World J Radiol* 2023; 15(12): 359-369
URL: https://www.wjgnet.com/1949-8470/full/v15/i12/359.htm
DOI: https://dx.doi.org/10.4329/wjr.v15.i12.359

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer in the United States, developing in about 4.3% of men and 4.0% of women. It is the second highest cause of cancer-related deaths in the United States, responsible for about 53200 deaths per year[1]. Likewise, CRC has also become the third most common cancer in China and is increasing in incidence in major countries such as Brazil and Russia[2]. Although early detection of CRC through established screening can greatly increase survival probability, resistance to the various invasive and noninvasive forms of screening persists[3-5]. This is reflected in the fact that up to 40% of CRC is diagnosed in the emergency department[6]. The rise in CRC incidence in adults younger than 55 also indicates a need for improved detection through non-screening methods[7-9]. Therefore, cross-sectional imaging remains important in early incidental diagnosis of CRC. However, up to 40% of the features of early CRC can be missed by radiologists when analyzing these scans[10-13]. This indicates a need for a "second-observer" to assist the busy radiologist in order to minimize false negatives which can result in reduced survival due to a delay in diagnosis[10].

Artificial intelligence (AI) has the potential to improve early disease detection, as shown by the recently approved algorithm for detecting intracranial hemorrhage on computed tomography (CT) and has been proposed for similar application in gastric cancer[14,15]. This model can be trained with the relatively low 39000 cases because of the low variation in brain anatomy and the simpler disease pattern on CT. CRC varies significantly in location and appearance because of the heterogeneity in anatomy and disease. The model training could be accomplished by using supervised

training (requires ground truth with potentially fewer training cases) rather than unsupervised training (uses more training cases without ground truth).

In supervised training, inputs with established ground-truth is important for training model. Although supervised training requires lower volume of data, the necessary volume of training data is quite large and would require significant amount of time to label the ground-truth by trained personnel[16-18]. Several potential methods for decreasing the time to establish the ground truth for supervised training have been evaluated. These methods include image level labeling, bounding boxes to localize the site of cancer, sparse labeling, and deploying an incompletely trained AI-model for first pass segmentation followed by human adjustment (AI-Init)[19,20]. Image level labeling and bounding box techniques consider different level of image information for localizing cancer and are expected to require significantly less human intervention. However, these may require more cases for training. For true supervised training, the level of annotation contains significantly more detail, which requires human interaction. This human interaction can be minimized with sparse annotation that skips slices during segmentation or having the rudimentarily trained AI algorithm perform the initial segmentation, which is subsequently modified by humans. In this report, we compared the improvement in amount of time spent on annotating the CRC between the last two techniques (sparse annotation by skip-slicing and AI-Init).

MATERIALS AND METHODS

CRC cases

The CRC cases were obtained from our respective cancer centers, which are tertiary and quaternary referral centers, between the years of 2012 through 2018 as well as cases from The Cancer Imaging Archive (TCGA-COAD) public domain images which were used to compare the outcomes of the techniques[21]. Our training and validating cases included 58 CRC cancer cases (27 males and 31 females) and 59 normal cases and consisted of CT scans of the abdomen and pelvis cases with a mixture of intravenous (IV) contrast enhanced and unenhanced studies. 51 of the cancer cases were used for training while 7 were retained for validation. The cases were retrieved from the picture archiving and communications system of the respective institutions and de-identified. The de-identified cases and their annotations were transferred between the research sites and the medical centers through the HIPPA compliant cloud server from Box.com. 20 of the 25 cases (8 males and 12 female) from TCGA-COAD were used to test the outcome of the training using the two different training techniques. The 5 excluded cases did not have clearly identifiable CRC on CT scans. The imaging stage of the 58 CRC cases and the 20 The Cancer Imaging Archive (TCIA) cases are listed in Table 1.

Cancer annotation

The location and slices of the cancer were identified on the CT images using ITK-SNAP (versions 3.6.0 and 3.8.0; www.itksnap.org)[22]. For annotation, the CT axial slices containing the CRC from our cancer center and from TCGA-COAD were identified and a contour outlining the edges of the cancer was drawn using the drawing tool. All the CT slices containing the tumor were segmented. At the time of training, to simulate sparse annotation, the skip-slice training would evenly skip one or two annotated slices among those containing the tumor from being used in training AI methods for every annotated tumor slice used for training. For AI-Init technique, the TCGA-COAD cases were initially segmented by the trained AI-model after training with the 51 CRC and 59 normal cases described in the previous section. This segmented model was then viewed with ITK-SNAP and adjusted to the ground truth established by human segmentation. The time required to fully adjust the contour and to eliminate false positives and to correct the false negatives was recorded for each TCGA-COAD case.

AI algorithm

The AI algorithm used in the project is a 2D U-Net, which is a convolutional neural network (CNN). The U-Net is a popular image segmentation algorithm for medical image segmentation tasks because it requires less training inputs than other techniques and is more robust with small training dataset[23]. In addition, recent research findings show that 2D U-Net has equivalent performance as that of 3D U-Net, but with lower computational requirement[24,25]. Inputs to the CNN consisted of 2D images with 512 × 512 pixels. The training dataset was augmented using the standard affine transformation with up to 30° rotation and up to 30% scale variation applied to the training patches. The CNN uses a 3 × 3 kernel and has 5 encoding layers containing 32, 64, 128, 256, and 512 filters and 5 decoding layers with each layer containing 512, 256, 128, 64, and 32 filters, respectively. Adam optimizer was used for training the model, and the model with the best validation accuracy was chosen. The network was trained using all the training cases, treating each image of a subject's study as a case, with 1 case per batch and trained with 200 epochs. The 7 validation cases were used to choose the optimal model parameters while the 20 TCGA-COAD cases were used to validate the final accuracy of the different techniques. For evaluating the effects of sparse-annotation, the AI model was trained with either all of the slices of annotation or evenly skipping either 1 or 2 slices of annotation for every slice used for training.

In order to determine how to improve the AI-Init technique in establishing the ground-truth, we also developed a simple ensemble model with each individual component of the ensemble being an independently trained 2D U-Net model with a random initiation. This was obtained to improve the specificity of the AI segmentation. To this end, we trained five randomly initialized U-Net models for voting-based model ensemble. Each of these five model is trained as described in the previous paragraph. The difference is that the final decision is based on the voter ensemble inference technique. For voting-based model ensemble, a voxel is labeled as tumor by the algorithm if and only if at least certain number of automatic segmentation models label the voxel as tumor. The U-Net models were also trained with the skip

Table 1 Clinical and protocol details of training and test cases

	Training cases (*n* = 58)	Test cases (*n* = 20)
AJCC stage		
Stage 1	0	5
Stage 2	15	5
Stage 3	14	7
Stage 4	29	3
T stage		
T1	0	2
T2	0	4
T3	21	13
T4	37	1
Location		
Right	39	17
Transverse	3	2
Left	16	1
CT slice thickness (mm)		
7	0	1
5	29	17
3-4	25	0
2 or less	4	2
Contrast		
IV+PO	27	18
IV	22	1
PO	4	1
None	5	0

AJCC: American Joint Commission on Cancer; CT: Computed tomography; IV: Intravenous; PO: Positive oral.

annotation technique (with skip 0, 1, and 2 slices). We then compared the accuracy and number of false positives per case to see how the ensemble method would influence the time required to adjust the segmentation results. For this approach, the model ensemble approach generated 3 unique segmentation algorithms (for each annotation technique) based on whether at least one, two, or three member(s) of the ensemble identified the same lesion. The segmentation was reviewed using ITK-SNAP.

Accuracy and sensitivity analysis

To analyze the algorithm's accuracy, the AI-generated segmentations were compared to ground-truth annotations, which were established as described previously. The AI model generated a DICOM image series for each case with the number of slices for each case ranging from 60-400. The cancer segmented by the AI model was compared to the annotated ground truth in 3D to determine the false positives, false negatives, and the Dice coefficient (DSC). A false positive was considered any segmentation created by the algorithm that did not overlap any part of the ground-truth segmentation. A false negative was considered as any image series with human annotated tumor that was not identified by the algorithm's segmentations. For true positive segmentation, DSC was visually estimated and categorized to be 0%-25%, 26%-50%, and > 50%. We obtained visual estimate as we do not have readily available software for full DSC calculation.

Time analysis for AI-Init and skip-slice annotation methods

In order to measure the amount of time saved by initiating annotation by the rudimentarily-trained AI model (as described earlier), we recorded the time required for annotating the CRC. The time required for initial, full annotation of CRC as well as the time required to adjust the AI produced model were acquired. For adjusting the AI-model, the obtained time included the time to adjust the boundary of the CRC and for erasing the false positives. We randomly selected 3 large, 3 medium, and 3 small CRC from the TCIA dataset and analyzed these times. The sizes of the CRC were

considered small, medium, or large if the lesion spanned ≤ 5 CT slices, 5-15 CT slices, and ≥ 15 CT slices, respectively. The median and average times were calculated. These same 9 cases were used for measuring the time needed to complete skip-slice annotation by annotating every other or every third slices of the tumor mass.

RESULTS

Comparison of training and TCIA datasets

The details of the location of the tumor and the scanning protocols from the training and TCIA testing datasets are listed in Table 1. The training dataset contains more higher stage cancers with 74 % of cases at stage 3 or 4 while the TCIA dataset has 50% of the cases at stage 3 or 4. All of the training cases is T3 or T4 in T-stage, while the TCIA dataset has 70% of the cases being T3 or T4. 67% of the training dataset has right-sided tumor (being defined as ascending colon) while the TCIA test dataset has 85% of the cases being right-sided. In terms of scanning protocol, 50% of the training dataset has 5 mm slice thickness while 90% of the TCIA testing dataset has slice thickness being 5 mm or more. In terms of contrast administration, the training dataset has 47% of the cases with IV and positive oral (PO) contrast while the TCIA dataset has 90% of the cases with both IV and PO contrast. 38% of the training dataset has just IV contrast whereas TCIA data has 5% of cases with IV contrast only.

Segmentations from skip-slice annotation trained AI-model

The AI-models generated by skip-slice annotation did not significantly alter the segmentation outcome of the AI-model. Figure 1 shows two separate cases segmented by the AI-models; although there is very subtle difference in the segmentations, the difference could not be detected by the measure that we chose (false positives per case, sensitivity, and DSC). For all three models, the sensitivity was 80% and the false positive lesions identified per case was 22. The DSC category distribution was 25% for 0-0.25, 60% for 0.26-0.5, and 15% for > 0.5.

Ensemble voting for decreasing false positives per case

Prior to obtaining the AI-initiated annotation, we aimed to minimize the number of false positives per case as the false positives could decrease the efficiency of this technique in establishing the ground-truth. To do this, we chose a simple voting-based ensemble method to reduce the number of false positives per case. When the number of votes required by the ensemble technique for determining tumor segmentation is increased, there is a corresponding drop in false positives per case while there is also a decrease in sensitivity, although the drop in false positives was much greater than the drop in sensitivity. The DSC distribution also shifts toward more cases being in 0 to 0.25 category. These data are shown in Tables 2 and 3. Figures 2 and 3 show an example of both agreement and disagreement between 1- and 2-voter models.

Time needed to adjust AI-Init segmentation and to complete skip-slide annotation

The models from the section above were used to generate the initial annotation of CRC which was then adjusted manually to fit the established ground truth. The amount of time required to modify these annotations to the ground truth was then recorded for 3 randomly selected cases from each of the large, medium, and small tumors. The complexity of these cases was determined as described in the methods section. The amount of time required to adjust these cases is listed in Table 4, along with the median and average. The measured time includes the time needed to remove the false positives as well as contouring the false negative lesions. The data show that AI-Init does decrease the time required to annotate the cases, although a statistical test of the distributions among the measured annotation time from the original, 1-voter, and 2-voter model using the Friedman's two-way analysis of variance by ranks did not yield statistical significance ($P = 0.121$). Some improvement is seen, primarily, with medium sized tumors.

For skip-slice annotation, the actual timed annotation revealed significant reduction in time needed to complete the annotation (Table 4). Although the reduction is not proportional, the differences are significant between full annotation and either skip-1 or skip-2 slice methods, using the Friedman's two-way analysis of variance by ranks. The *P*-values for univariate analysis between fully-annotated and skip-1, fully-annotated and skip 2, and skip-1 and skip-2 annotation style are 0.034, < 0.001, and 0.034. When using multivariate analysis, the same *P*-values are 0.102, < 0.001, and 0.102. This suggests that skipping slices can reduce the labor necessary for establishing the ground-truth for supervised or semi-supervised training of AI models, and in multivariate analysis, the time different is statistically significant when higher number of slices are skipped.

DISCUSSION

Our results provide a direct comparison of annotation techniques for supervised training of AI models as second observer for detecting CRC. Supervised AI-model training by skipping-slices of CRC did not appreciably influence the outcome of segmentation. There were very subtle visual differences, but these were not detectable with the measures used. No significant segmentation difference could be detected when skipping up to 2 slices. For AI-initiated annotation, the model does not improve the time spent annotating large and small cancers, but does show some improvement for medium sized tumors. These methods allow for time-reduction in annotating the ground truth so supervised training of AI-models could be more efficient and allow greater participation by busy radiologists.

Table 2 Sensitivity and false positives/case for ensemble technique

	Single voter	2 voter	3 voter
Sensitivity	0.8	0.6	0.3
False positives/case	21.95	7.55	3.7

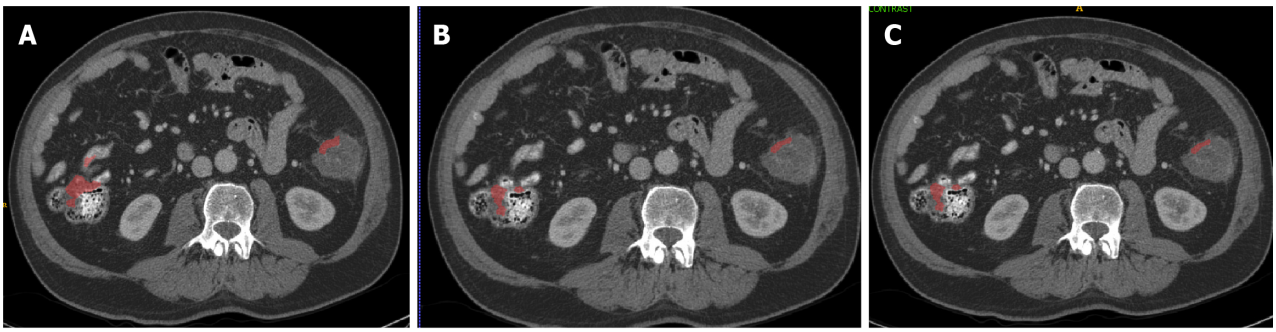
Table 3 Dice coefficient distribution for ensemble technique

Percentage of cases	Estimated dice coefficient			
	0	0-0.25	0.25-0.5	> 0.5
Single voter	20	5	60	15
2 voter	40	35	20	5
3 voter	70	15	10	5

Table 4 Amount of time needed to annotate the tumor

Lesion size	Annotation time based on technique (Min:Sec ± min)				
	Manual (*n* = 3 each)	AI-single voter (*n* = 3 each)	AI-2-voter (*n* = 3 each)	Skip-1 (*n* = 3 each)	Skip-2 (*n* = 3 each)
Large	22:09 ± 0.18	21:00 ± 0.23	20:29 ± 0.22	8:58 ± 1.22	5:34 ± 1.19
Medium	15:06 ± 0.4	10:37 ± 0.25	9:13 ± 0.15	4:58 ± 2.57	1:14 ± 1.38
Small	5:54 ± 0.07	6:26 ± 0.03	5:44 ± 0.02	2:23 ± 0.14	1:24 ± 0.28

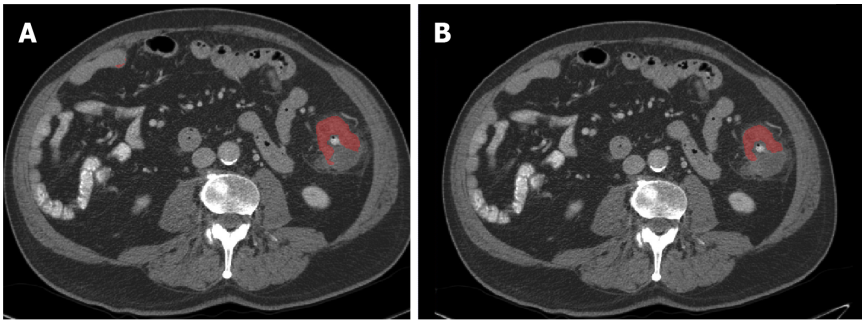
AI: Artificial intelligence.



DOI: 10.4329/wjr.v15.i12.359

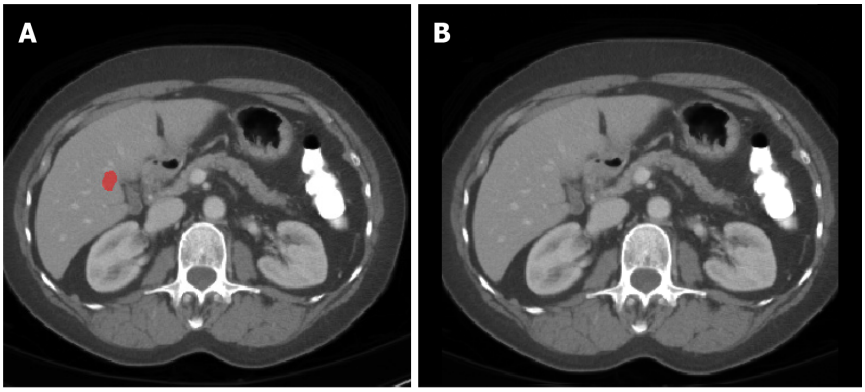
Figure 1 Artificial intelligence segmentation by models with skipped slice training. A-C: Artificial intelligence (AI) segmented lesion by model trained without skipping slices (A), with skipping 1 slice (B), and with skipping 2 slices (C). There is slight difference in the segmentation, but insufficient to modify the Dice coefficient. The cancer is in the descending colon, only a small portion of which was segmented by AI model. The slightly larger false positive lesion may be due to slightly different slice level.

Skip-slice annotation is similar to a sparse-annotation technique, which has been explored in the literature. This technique was tested in confocal images of Xenopus kidney segmentation[19]. Cicek *et al*[19] showed that the DSC equivalent (intersection of union) improved with increasing slices in each axis, starting with one slice in each direction. Increasing the annotated slices was equivalent to increasing the number of ground-truth pixels based on their study. The authors achieved their annotation goal using up to 9% of all the available pixels as ground truth for training. With this training, the network achieved segmentation with 85% overlap with the ground truth. Our data also support this finding that not all ground truth needs to be presented to an AI-algorithm to train the algorithm properly. The difference between Cicek *et al*[19] and our study is that they began with the minimal number of slices while we evaluated from the maximum number of slices. Increasing from the minimum showed that minimalist approach may underfit the algorithm, while maximal approach may over-fit the network. The minimum necessary amount of established ground truth pixel for optimal network training is yet to be identified so that the amount of human effort in establishing ground truth can be minimized while maintaining optimal AI model training.



DOI: 10.4329/wjr.v15.i12.359 **Copyright** ©The Author(s) 2023.

Figure 2 Examples of lesion agreement by 1- and 2-voter ensemble technique. A and B: 1- (A) and 2- (B) voter(s) model agreeing on the same tumor mass, although 2-voters mark less of the mass.



DOI: 10.4329/wjr.v15.i12.359 **Copyright** ©The Author(s) 2023.

Figure 3 Example of lesion disagreement by 1- and 2-voter ensemble technique. A and B: 1- (A) voter model marks a false positive in the liver which is rejected by 2- (B) voter model.

The literature reports multiple techniques for minimizing false positives, particularly regarding pulmonary nodule reduction which can be categorized into ones that use single modes of information (imaging only) or multimodal technique (combines imaging with clinical information). Jin *et al*[26] constructed a false-positive reduction algorithm for pulmonary nodule detection. This is constructed as a separate algorithm that could serve as add on to a nodule detector. They combined several methods to avoid the traps that cause false-positives. First, they deployed a 3D residual CNN which minimizes the effect of diminishing gradient in the stochastic gradient descent algorithm during training, so as to avoid local minima that may trap the algorithm. They also combined spatial pooling and cropping which provides multi-scale contextual information to improve the learning process. A similar technique uses multiscale contextual information where variable amount of the pulmonary nodule and surrounding normal lung is included for training[27,28]. The multiple levels of information are integrated and significantly lower the false positives[27,28]. Lastly, online hard sample selection training was chosen to maximize training of hard-to-discern examples so that the network can learn from its own failures. This technique replaces a portion of correctly detected training cases with ones that were previously missed so that the network can learn the features of the missed cases to improve its outcome[29].

The multimodal technique is a broad category of AI technique where different aspects of a patient's clinical information are integrated to improve the classification and prediction algorithm. The additional information restricts the bias and variance of the model to improve the accuracy of the outcome[30]. For our algorithm, we employed the simulated multimodal technique with ensemble voting by integrating information from different instances of the AI model. This is similar to selecting 5 different models from the same model space to restrict bias and variance[30]. By generating 5 models from the base CNN technique, we chose a lesion to be cancer only if 2 or 3 of the 5 models agreed on a pixel being cancerous. This allowed us to dramatically decrease the number of false positives per case. This, however, also decreased sensitivity of the model. This trade-off is also seen with ensemble techniques trained with clinical data in predicting diabetic retinopathy[31]. Other studies have shown that the information contains different degree of relevance [32,33]. In the study by Boehm *et al*[32], applying all available information regarding a patient (clinical, genetic, histological, and radiological) resulted in less accurate outcome than one that deployed a limited dataset (genetic, histological, and radiological). Likewise the study by Iseke *et al*[33] which used both clinical and imaging information to predict hepatocellular carcinoma recurrence after treatment did not achieve a better prediction than using imaging alone. Multimodal AI can provide better outcomes, but only with the appropriate dataset; overloading the AI system with lower relevance data may over-fit the system to a less than optimal parameter space.

The findings that skip-slice annotation may reduce the time required for establishing the ground-truth for AI model training can significantly impact the development of AI models in imaging research. Annotating the ground truth requires trained personnel capable of identifying normal and abnormal structures on CT images, who are typically physicians or physicians in training. As we have shown in Table 4, full annotation of a case will require anywhere from 5-20 min, which can be reduced significantly with skip-slice annotation. This will significantly reduce the time of the highly trained personnel, who are involved in busy clinical work. Minimizing the time spent in establishing the ground-truth should theoretically improve participation of these highly trained personnel in assisting AI research in medical imaging.

There are several limitations to the present study. The first is the small size of the training and testing dataset. These training and testing dataset is also unbalanced and unmatched in both stage of the disease and the scanning protocol. Table 1 showed that the training dataset consists of higher stage disease than the testing dataset which may limit detection of the earlier stage disease in the testing dataset. In addition, the training dataset also has lower proportion of the cases containing both IV and PO contrast and has thinner slice images. It is uncertain if the blur from the thicker slices may influence the decision of the model, but the trained model have been exposed to thicker slices with IV and PO contrast. It will be interesting to evaluate the dilutive effect on model performance when the model is trained with a broader range of protocols and stages of the disease. The current model with limited training dataset is not generalizable, but it does show the potential of the second observer with better trained model[34]. Another limitation of the study is the lack of full software for calculating the precision, F1 score, and DSC of the model outcome. The sensitivity provided in the present study is equivalent to the recall measure of the model.

CONCLUSION

In comparing the different techniques for reducing annotation time to establish the ground truth, we developed a U-NET model in detecting CRC. This pilot model has the potential to serve as a second observer with further research. In order to accelerate AI second observer training, we compared different techniques of annotation in minimizing this data preparation work. Our results showed that skip slice annotation may lead to the most time reduction as there was minimal effect on model outcome when slices are skipped, leading to proportional decrease in time needed to annotate. Although AI-initiated segmentation may lead to reduced annotation time, it tends to reduce time for medium sized lesion while large complex and small lesions do not benefit. At this time, skipping slices may result in the most time efficient method for annotating cancer on training images.

ARTICLE HIGHLIGHTS

Research background

Up to 40% of colorectal cancer (CRC) goes undetected on initial computed tomography (CT) scan performed in either the emergency department or outpatient imaging setting. This delay in diagnosis significantly impacts the overall survival of the patients. The ultimate goal is to develop an artificial intelligence (AI)-based second observer for clinical integration so as to improve the clinical diagnosis of CRC on CT studies.

Research motivation

The development of deep learning has shown that AI can potentially serve as a second observer to assist busy radiologist at a reasonable cost, as second reader has been shown in past research to improve imaging diagnosis. However, to develop an AI second observer, large number of training cases with annotated ground truth is required necessitating significant time commitment on the part of the research radiologists.

Research objectives

Our main objective in this research is to compare skip-slice annotation with AI-initiated annotation in time savings for annotating the ground truth for training dataset preparation. Saving annotation time will help improve the efficiency in dataset preparation. Our secondary objective was to evaluate whether ensemble technique could help improve false positive rate for AI-initiated annotation technique. Decreasing false positives per case will make the model more acceptable by clinical radiologist.

Research methods

The dataset was manually annotated for the entire tumor as well as skipping annotation by one or two slices was measured; 9 total cases were randomly selected to measure the time required to annotate these tumors. These datasets were used to train 2D U-Net model with 5 encoding and 5 decoding layers, using the Adam optimizer. The model accuracy consisting of sensitivity, Dice coefficient estimate, and false positive per case were used to evaluate the model accuracy. The rudimentary AI model was also used to annotate the ground truth; the times required to adjust the annotation for the 9 cases from manually annotation were also measured.

Research results

We found that the model trained on skip-slice annotation did not have significant difference in tumor segmentation as a

fully annotated dataset and which is statistically significant, thus showing that skip slice annotation can reduce the data preparation time. Although AI-initiated annotation also reduces time, the difference was not statistically significant. Ensemble technique is shown to reduce false positive per case, but at decreased sensitivity.

Research conclusions

This study proposes that skip-slice annotation can improve the efficiency in data preparation for AI model training. The significance is that it will reduce the time commitment of highly trained medical personnel in participating in AI medical imaging research.

Research perspectives

The future direction of the present research is that this should improve the efficiency in training dataset development given the decreased annotation time.

ACKNOWLEDGEMENTS

The authors would like to thank Gene A Hoyer and Judith C Hoyer for their generous support in funding the publication of this manuscript.

FOOTNOTES

Author contributions: Grudza M and Kodibagkar V contributed to the data analysis and initial write up of the manuscript; Salinel B, Zeien S, and Murphy M contributed to the ground truth of the training and testing dataset; Adkins J and Jensen CT contributed to the data curating from MD Anderson; Syeda-Mahmood T and Wang HZ contributed to the AI Model development and training; Bay C contributed to the statistical analysis of the data; Koo P, Dragovich T, Choti MA, and Kundranda M contributed to the Banner MD Anderson data collection and manuscript revision; Chang J conceived and oversaw the entire project and the manuscript writeup.

Institutional review board statement: The study was reviewed and approved by the Banner MD Anderson Cancer Center IRB.

Informed consent statement: The study was approved by Banner MD Anderson Cancer Center IRB with exemption for individual consent due to retrospective nature of the data collection.

Conflict-of-interest statement: All the authors report no relevant conflicts of interest for this article.

Data sharing statement: Dataset can be available by contacting the corresponding author at john.chang@bannerhealth.com.

STROBE statement: The authors have read the STROBE Statement-checklist of items, and the manuscript was prepared and revised according to the STROBE Statement-checklist of items.

Country/Territory of origin: United States

ORCID number: Tomislav Dragovich 0000-0002-1569-2172; Madappa Kundranda 0000-0002-9237-2666; John Chang 0000-0003-2896-1586.

S-Editor: Wang JJ
L-Editor: A
P-Editor: Zhao S

REFERENCES

1 **American Cancer Society**. Key Statistics for Colorectal Cancer. [cited 15 July 2023]. Available from: https://www.cancer.org/cancer/types/colon-rectal-cancer/about/key-statistics.html
2 **Xie Y**, Shi L, He X, Luo Y. Gastrointestinal cancers in China, the USA, and Europe. *Gastroenterol Rep (Oxf)* 2021; **9**: 91-104 [PMID: 34026216 DOI: 10.1093/gastro/goab010]
3 **Akram A**, Juang D, Bustamante R, Liu L, Earles A, Ho SB, Wang-Rodriguez J, Allison JE, Gupta S. Replacing the Guaiac Fecal Occult Blood Test With the Fecal Immunochemical Test Increases Proportion of Individuals Screened in a Large Healthcare Setting. *Clin Gastroenterol Hepatol* 2017; **15**: 1265-1270.e1 [PMID: 28167157 DOI: 10.1016/j.cgh.2017.01.025]
4 **Weiser E**, Parks PD, Swartz RK, Thomme JV, Lavin PT, Limburg P, Berger BM. Cross-sectional adherence with the multi-target stool DNA test for colorectal cancer screening: Real-world data from a large cohort of older adults. *J Med Screen* 2021; **28**: 18-24 [PMID: 32054393 DOI:

10.1177/0969141320903756]
5 **Gupta S**. Screening for Colorectal Cancer. *Hematol Oncol Clin North Am* 2022; **36**: 393-414 [PMID: 35501176 DOI: 10.1016/j.hoc.2022.02.001]
6 **Esteva M**, Ruidíaz M, Sánchez MA, Pértega S, Pita-Fernández S, Macià F, Posso M, González-Luján L, Boscá-Wats MM, Leiva A, Ripoll J; DECCIRE GROUP. Emergency presentation of colorectal patients in Spain. *PLoS One* 2018; **13**: e0203556 [PMID: 30273339 DOI: 10.1371/journal.pone.0203556]
7 **Dharwadkar P**, Zaki TA, Murphy CC. Colorectal Cancer in Younger Adults. *Hematol Oncol Clin North Am* 2022; **36**: 449-470 [PMID: 35577711 DOI: 10.1016/j.hoc.2022.02.005]
8 **Fisher DA**, Saoud L, Finney Rutten LJ, Ozbay AB, Brooks D, Limburg PJ. Lowering the colorectal cancer screening age improves predicted outcomes in a microsimulation model. *Curr Med Res Opin* 2021; **37**: 1005-1010 [PMID: 33769894 DOI: 10.1080/03007995.2021.1908244]
9 **Wu Y**, Jiao N, Zhu R, Zhang Y, Wu D, Wang AJ, Fang S, Tao L, Li Y, Cheng S, He X, Lan P, Tian C, Liu NN, Zhu L. Identification of microbial markers across populations in early detection of colorectal cancer. *Nat Commun* 2021; **12**: 3063 [PMID: 34031391 DOI: 10.1038/s41467-021-23265-y]
10 **Rodriguez R**, Perkins B, Park PY, Koo PJ, Kundranda M, Chang JC. Detecting early colorectal cancer on routine CT scan of the abdomen and pelvis can improve patient's 5-year survival. *Arch Biomed Clin Res* 2019; **1**: 1-6 [DOI: 10.15761/ABCR.1000102]
11 **Mangat S**, Kozoriz MG, Bicknell S, Spielmann A. The Accuracy of Colorectal Cancer Detection by Computed Tomography in the Unprepared Large Bowel in a Community-Based Hospital. *Can Assoc Radiol J* 2018; **69**: 92-96 [PMID: 29458958 DOI: 10.1016/j.carj.2017.12.005]
12 **Balthazar EJ**, Megibow AJ, Hulnick D, Naidich DP. Carcinoma of the colon: detection and preoperative staging by CT. *AJR Am J Roentgenol* 1988; **150**: 301-306 [PMID: 3257314 DOI: 10.2214/ajr.150.2.301]
13 **Johnson CD**, Flicek KT, Mead-Harvey C, Quillen JK. Strategies for improving colorectal cancer detection with routine computed tomography. *Abdom Radiol (NY)* 2023; **48**: 1891-1899 [PMID: 36961532 DOI: 10.1007/s00261-023-03884-3]
14 **Arbabshirani MR**, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, Moore GJ. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med* 2018; **1**: 9 [PMID: 31304294 DOI: 10.1038/s41746-017-0015-z]
15 **Cao R**, Tang L, Fang M, Zhong L, Wang S, Gong L, Li J, Dong D, Tian J. Artificial intelligence in gastric cancer: applications and challenges. *Gastroenterol Rep (Oxf)* 2022; **10**: goac064 [PMID: 36457374 DOI: 10.1093/gastro/goac064]
16 **Bangert P**, Moon H, Woo JO, Didari S, Hao H. Active Learning Performance in Labeling Radiology Images Is 90% Effective. *Front Radiol* 2021; **1**: 748968 [PMID: 37492167 DOI: 10.3389/fradi.2021.748968]
17 **Sermesant M**, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardiovascular imaging. *Nat Rev Cardiol* 2021; **18**: 600-609 [PMID: 33712806 DOI: 10.1038/s41569-021-00527-2]
18 **Willemink MJ**, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, Folio LR, Summers RM, Rubin DL, Lungren MP. Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020; **295**: 4-15 [PMID: 32068507 DOI: 10.1148/radiol.2020192224]
19 **Cicek O**, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. 2016 Preprint. Available from: arXiv:1606.06650 [DOI: 10.48550/arXiv.1606.06650]
20 **Kervadec H**, Dolz J, Wang S, Granger E, Ayed IB. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision. 2004 Preprint. Available from: arXiv:2004.06816v1 [DOI: 10.48550/arXiv.2004.06816]
21 **The Cancer Imaging Archive**. The Cancer Genome Atlas Colon Adenocarcinoma Collection (TCGA-COAD). [cited 15 July 2023]. Available from: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=16712033
22 **Yushkevich PA**, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006; **31**: 1116-1128 [PMID: 16545965 DOI: 10.1016/j.neuroimage.2006.01.015]
23 **Ronneberger O**, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. 2015 Preprint. Available from: arXiv:1505.04597v1 [DOI: 10.48550/arXiv.1505.04597]
24 **Srikrishna M**, Heckemann RA, Pereira JB, Volpe G, Zettergren A, Kern S, Westman E, Skoog I, Schöll M. Comparison of Two-Dimensional- and Three-Dimensional-Based U-Net Architectures for Brain Tissue Classification in One-Dimensional Brain CT. *Front Comput Neurosci* 2021; **15**: 785244 [PMID: 35082608 DOI: 10.3389/fncom.2021.785244]
25 **Zettler N**, Mastmeyer A. Comparison of 2D vs. 3D U-Net Organ Segmentation in abdominal 3D CT images. 2021 Preprint. Available from: arXiv:2107.04062v1 [DOI: 10.48550/arXiv.2107.04062]
26 **Jin H**, Li Z, Tong R, Lin L. A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. *Med Phys* 2018; **45**: 2097-2107 [PMID: 29500816 DOI: 10.1002/mp.12846]
27 **Dou Q**, Chen H, Yu L, Qin J, Heng PA. Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Trans Biomed Eng* 2017; **64**: 1558-1567 [PMID: 28113302 DOI: 10.1109/TBME.2016.2613502]
28 **Kim BC**, Yoon JS, Choi JS, Suk HI. Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection. *Neural Netw* 2019; **115**: 1-10 [PMID: 30909118 DOI: 10.1016/j.neunet.2019.03.003]
29 **Shrivastava A**, Gupta A, Girschick R. Training region-based object detectors with online hard example mining. 2016 Preprint. Available from: arXiv:1604.03540v1 [DOI: 10.48550/arXiv.1604.03540]
30 **Seni G**, Elder JF. Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. In: Synthesis Lectures on Data Mining and Knowledge Discovery. Germany: Springer, 2010
31 **Yang T**, Zhang L, Yi L, Feng H, Li S, Chen H, Zhu J, Zhao J, Zeng Y, Liu H. Ensemble Learning Models Based on Noninvasive Features for Type 2 Diabetes Screening: Model Development and Validation. *JMIR Med Inform* 2020; **8**: e15431 [PMID: 32554386 DOI: 10.2196/15431]
32 **Boehm KM**, Aherne EA, Ellenson L, Nikolovski I, Alghamdi M, Vázquez-García I, Zamarin D, Long Roche K, Liu Y, Patel D, Aukerman A, Pasha A, Rose D, Selenica P, Causa Andrieu PI, Fong C, Capanu M, Reis-Filho JS, Vanguri R, Veeraraghavan H, Gangai N, Sosa R, Leung S, McPherson A, Gao J; MSK MIND Consortium, Lakhman Y, Shah SP. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat Cancer* 2022; **3**: 723-733 [PMID: 35764743 DOI: 10.1038/s43018-022-00388-9]
33 **Iseke S**, Zeevi T, Kucukkaya AS, Raju R, Gross M, Haider SP, Petukhova-Greenstein A, Kuhn TN, Lin M, Nowak M, Cooper K, Thomas E, Weber MA, Madoff DC, Staib L, Batra R, Chapiro J. Machine Learning Models for Prediction of Posttreatment Recurrence in Early-Stage Hepatocellular Carcinoma Using Pretreatment Clinical and MRI Features: A Proof-of-Concept Study. *AJR Am J Roentgenol* 2023; **220**: 245-255 [PMID: 35975886 DOI: 10.2214/AJR.22.28077]

34 **Korfiatis P**, Suman G, Patnam NG, Trivedi KH, Karbhari A, Mukherjee S, Cook C, Klug JR, Patra A, Khasawneh H, Rajamohan N, Fletcher JG, Truty MJ, Majumder S, Bolan CW, Sandrasegaran K, Chari ST, Goenka AH. Automated Artificial Intelligence Model Trained on a Large Data Set Can Detect Pancreas Cancer on Diagnostic Computed Tomography Scans As Well As Visually Occult Preinvasive Cancer on Prediagnostic Computed Tomography Scans. *Gastroenterology* 2023; **165**: 1533-1546.e4 [PMID: 37657758 DOI: 10.1053/j.gastro.2023.08.034]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com