Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and validation of an artificial intelligence-based virtual indigo carmine chromoendoscopy model: a multicenter study

Xianglei Yuan
West China Hospital of Sichuan University
Xianhui Zeng
West China School of Public Health and West China Fourth Hospital
Ruide Liu
West China Hospital of Sichuan University
Yuan Gao
West China Hospital of Sichuan University
Wanhong Zhang
Cangxi People's Hospital
Xiaogang Bi
Zigong Fourth People's Hospital
Tao Zhang
Nanchong Central Hospital
Linling Liu
Zigong Fourth People's Hospital
Luli Hu
Nanchong Central Hospital
Congxin Liang
Southwest Jiaotong University
Yuhang Zhang
West China Hospital of Sichuan University
Liuxiang Chen
West China Hospital of Sichuan University
Qi Luo
West China Hospital of Sichuan University
Nuoya Zhou
West China Hospital of Sichuan University
Pinghua Wen
West China Hospital of Sichuan University
Huan Jiang

West China Hospital of Sichuan University

Bing Hu

West China Hospital of Sichuan University

Weili Jiang

Southwest Jiaotong University

Yi Mou

1256546390@qq.com

West China Hospital of Sichuan University

Research Article

Keywords: Virtual indigo carmine, artificial intelligence, chromoendoscopy, generative adversarial network

Posted Date: September 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10454340/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Indigo carmine chromoendoscopy (ICC) is an important technique for detecting early gastric and colorectal cancers. However, it is time-consuming and highly dependent on endoscopists experience, limiting its standardized application in clinical practice. This study aimed to develop and systematically evaluate an artificial intelligence (AI)-based virtual ICC model that transforms white light images into virtual ICC images.

Methods: This study included 9,174 ICC and 8,403 white light images from 960 cases across five hospitals for model training and testing. A structure-guided multi-step virtual staining generation model based on CycleGAN was developed. Objective metrics for comparing different models included Fréchet Inception Distance (FID), Colorfulness Score (CF) with its differential metric ΔCF, and Structural Similarity Index (SSIM). Subjective evaluation was conducted by 11 endoscopists (5 seniors, 6 novices) assessing authenticity, contrast, lesion boundary clarity, and microstructural visibility.

Results: Our model achieved superior structural fidelity, color representation, and lower FID than comparative models. The accuracy of distinguishing virtual from real images was 78.1% (95% CI: 76.2%–79.9%). No significant differences were observed between virtual and real images in overall contrast (3.85±0.90 vs. 3.92±0.95, *P*=0.059) and lesion boundary clarity (3.72±0.96 vs. 3.76±0.99, *P*=0.157). But virtual images scored lower in microstructural visibility (3.35±0.96 vs. 3.61±1.03, *P*<0.001). Interobserver agreement for virtual images was moderate-to-good (intraclass correlation coefficient>0.5) and comparable to that for real images. No significant interaction effects were observed between staining modality (virtual vs. real) and endoscopists in terms of overall contrast and lesion boundary clarity (*P*>0.05).

Conclusions: The AI-based virtual ICC model achieved clinically acceptable performance in overall contrast and lesion boundary delineation, with good stability and generalizability. It might reduce reliance on conventional dye-based procedures and improve early gastric and colorectal cancers detection, though prospective real-world validation remains necessary.

Introduction

Gastric cancer and colorectal cancer are major gastrointestinal malignancies that impose a substantial global health burden. Because early-stage symptoms are often non-specific or absent, most patients are diagnosed at advanced stages, resulting in a 5-year survival rate of less than 20% and considerable socioeconomic costs[1, 2]. Therefore, early diagnosis and treatment are critical to enhancing patient prognosis and alleviating the associated disease burden.

Digestive endoscopy is the primary approach for detecting early gastric and colorectal cancers, including white-light imaging (WLI), electronic chromoendoscopy, and indigo carmine chromoendoscopy (ICC). WLI serves as the fundamental modality. Electronic chromoendoscopy offers high diagnostic sensitivity for early lesions; however, its diagnostic system is complex and costly[3, 4]. ICC, performed by spraying

indigo carmine onto the gastrointestinal mucosa via a catheter during WLI, exploits its deposition in mucosal depressions to enhance contrast between lesions and surrounding normal tissue. This morphological enhancement facilitates accurate delineation of lesion margins and gross characterization, providing essential guidance for targeted biopsy and therapeutic decisions, thereby improving detection rates and diagnostic accuracy[5, 6]. Given the uneven economic development across regions in China, many resource-limited facilities lack access to electronic chromoendoscopy, whereas ICC is widely adopted due to its low cost and favorable safety profile, as the dye is not absorbed by the mucosa. Thus, emphasizing ICC for early cancer diagnosis is clinically warranted. Nevertheless, ICC has inherent limitations: the staining procedure is time-consuming, potentially prolonging examination time and increasing anesthesia-related risks; moreover, staining quality is highly operator-dependent, leading to inconsistent quality and compromised lesion detection.

Recent advances in artificial intelligence (AI) have enabled broad applications in digestive endoscopy[7–10]. Through deep learning, AI can process WLI images to produce virtual images that mimic ICC features, thereby simulating expert-level chromoendoscopy without physical dye application. This technique preserves the diagnostic benefits of chromoendoscopy, while also simplifying clinical procedures, diminishing the reliance on operator expertise, and obviating potential risks associated with dye use. Virtual staining techniques based on generative adversarial networks (GANs)[11] have shown promise in medical imaging, particularly in virtual cell staining and medical image synthesis. Liu et al.[12] employed GAN to convert routine hematoxylin and eosin-stained images into virtual SOX10-stained images for melanocyte identification, offering a novel tool for melanoma diagnosis. Wölflein et al.[13] used GAN to virtually stain Hoechst images into CD3 and CD8 counterparts, enabling T-cell subtyping in clear cell renal carcinoma without physical staining. Lyu et al.[14] developed a GAN-based non-contrast CT angiography model for the aorta and carotid arteries, generating images comparable to true contrast-enhanced CT without contrast injection, thereby avoiding associated risks while aiding diagnosis.

To date, investigations into GAN-based virtual ICC for digestive endoscopy have indicated potential feasibility[15–19], yet notable shortcomings remain. These include a predominance of single-center designs with limited sample sizes, reliance on unmodified GAN or CycleGAN architectures unsuitable for endoscopic image characteristics, and the absence of rigorous, quantitative assessment criteria for virtual ICC image quality, with evaluations often based on small samples and narrow metrics. Therefore, this study aims to develop a dedicated virtual ICC model and systematically validate its generated images through multi-dimensional analyses.

Methods

Preparation of datasets

This multicenter retrospective study collected paired ICC and corresponding WLI images from five hospitals: West China Hospital, Sichuan University (WCHSCU); West China Fourth Hospital, Sichuan

University (WCFHSCU); Zigong Fourth People's Hospital (ZFPH); Cangxi People's Hospital (CPH); and Nanchong Central Hospital (NCH). Only images with adequate brightness and clear visualization were included.

For datasets partitioning, images from WCHSCU were chronologically allocated, with those acquired between July 2021 and May 2024 designated as the training set and those obtained from June to December 2024 as the test set. Images from WCFHSCU and ZFPH were incorporated into the training set, whereas images from CPH and NCH were assigned to the test set. Within the training set, images were randomly stratified into training and validation subsets at an 8:2 ratio, with strict patient-level separation to prevent data leakage. One virtual and one real ICC image were randomly selected per case from each hospital comprising an independent subjective evaluation set.

The study protocol was approved by the Biomedical Research Ethics Committee of WCHSCU [Approval No. 2024 Review (1337)], with informed consent waived for retrospective data.

Development of the AI model

To address the key clinical requirement for virtual ICC that tissue structures remain unchanged while staining appearance is controllably transformed, we proposed a structure-guided multi-step generation model based on an improved CycleGAN. Unlike the standard CycleGAN, which performs a direct one-step transformation, our model translates WLI to virtual images through a gradual, step-by-step process. In this way, the image progressively moves toward the target staining appearance while preserving essential anatomical details, such as glandular patterns, vessel paths, and lesion boundaries. Detailed descriptions of the model architecture and training procedures are available in the Supplementary Materials.

Validation of the AI model

A comparative study was conducted to evaluate the proposed model against representative algorithms from three categories: image colorization (DDColor)[20], style transfer (CycleGAN[21] and UNSB[22]), and diffusion probabilistic models (ILVR[23] and SD[24]). Quantitative assessments were based on four metrics, including Fréchet Inception Distance (FID), Colorfulness Score (CF) with its differential metric ΔCF, and Structural Similarity Index (SSIM). The ΔCF metric was defined as the absolute CF difference between the virtual and corresponding real ICC images. Detailed descriptions of the other metrics are provided in the Supplementary Materials.

A subjective evaluation of the virtual ICC images generated by our model was conducted by 11 endoscopists from five hospitals. The evaluation comprised two steps: first, distinguishing virtual from real images; second, rating image quality on a 5-point Likert scale (excellent, good, fair, poor, very poor) across three dimensions, namely overall contrast for lesion detectability, lesion boundary clarity, and microstructure visualization. Endoscopists were divided into senior (n = 5, > 8 years experience, > 5,000

gastroscopies) and novice (n = 6, < 1 year, < 500 procedures) groups. Metrics for ICC images included sensitivity, specificity, and accuracy, defined as the proportions of virtual images correctly identified, real images correctly identified, and overall correct classifications, respectively. Mean scores for virtual versus real images were compared across the three dimensions. Inter-rater agreement was assessed using the intraclass correlation coefficient, with < 0.5 indicating poor agreement, 0.5–0.75 moderate agreement, 0.76–0.9 good agreement, and > 0.9 excellent agreement. An ordinal logistic mixed-effects model was used to examine the interaction between staining method (virtual vs. real) and endoscopist experience.

Statistical analysis

Endoscopists' scores were expressed as mean ± standard deviation and median with interquartile range. Comparisons between groups were made using the Wilcoxon rank-sum test. The 95% confidence intervals (CI) were calculated using the Clopper-Pearson exact method. A two-sided P-value of < 0.05 was considered statistically significant. Statistical analysis was performed using Python 3.11 (Facebook, USA) and SPSS 26.0 (IBM, USA).

Results

Datasets information

960 cases with 9,174 ICC images and 8,403 WLI images were collected from the five hospitals. The training set comprised 765 cases (6,910 ICC, 6,451 WLI), while the test set comprised 195 cases (2,264 ICC, 1,952 WLI). The subjective evaluation set included 195 virtual and 195 real ICC images. Detailed datasets information is summarized in Table 1.

Table 1
Distribution of the Datasets.

Datasets	Hospitals	Times	Cases	ICC	WLI
Training	WCHSCU	2021.07-2024.05	606	5096	4740
	WCFHSCU	2023.01-2024-11	92	1547	1519
	ZFPH	2021.01-2024.03	67	267	192
Test	WCHSCU	2024.06-2024.12	61	547	580
	CPH	2018.06-2024.07	74	1210	927
	NCH	2020.01-2024.04	60	507	445
All	-	-	960	9174	8403

ICC, indigo carmine chromoendoscopy; WLI, white-light imaging; WCHSCU, West China Hospital, Sichuan University; WCFHSCU, West China Fourth Hospital, Sichuan University; ZFPH, Zigong Fourth People's Hospital; CPH, Cangxi People's Hospital; NCH, Nanchong Central Hospital.

Comparative results with representative algorithms (Objective evaluation)

As shown in Table 2, DDColor was outperformed by our model in FID and ΔCF, indicating that the virtual images generated by our model more closely approximate real images in both feature space distribution and color expressiveness. Regarding structural consistency, a slightly lower SSIM was obtained by our model compared with DDColor, suggesting a marginally reduced capacity for preserving tissue architecture during virtual staining, though the overall level remained high.

Table 2
Comparative results with representative algorithms.

Comparison	FID↓	CF↑	ΔCF↓	SSIM↑
DDColor	81.37	40.98	8.69	0.229
CycleGAN	83.87	35.90	3.61	0.167
UNSB	78.69	26.07	6.22	0.168
ILVR	155.27	59.69	27.40	0.150
SD	99.87	58.78	26.49	0.175
Ours	62.12	59.75	6.26	0.176
FID, Fréchet Inception Distance; CF, Colorfulness Score; SSIM, Structural Similarity Index.				

CycleGAN and UNSB were surpassed by our model across FID, CF, and SSIM. Specifically, the lowest FID (62.12) was achieved by our model, indicating closer feature-space proximity to real images and higher fidelity in overall style and content. The highest CF (59.75) was also attained by our model, reflecting superior color richness. In terms of SSIM, the structural information of input images was completely preserved by our model, outperforming both comparators. Although a higher ΔCF was yielded by our model than by CycleGAN and UNSB, more balanced overall performance was demonstrated by our model.

Compared with ILVR and SD, the best results across all four metrics were achieved by our model, demonstrating that the generated virtual images more closely resemble real images in feature space distribution, color richness, and color expressiveness, while the structural information of input images was fully preserved, with superior structural fidelity. Results stratified by hospital are available in the Supplementary Materials.

Comparative results with representative algorithms (Subjective Evaluation)

Figure 1 presents the virtual ICC images generated by different algorithms on the same set of WLI images.The virtual images generated by DDColor exhibited overly faint staining with insufficient color contrast, failing to adequately reproduce the visual characteristics of real images. Superior staining effects were achieved by CycleGAN and UNSB, compared with the DDColor. However, inconsistent performance was observed across samples: staining quality was close to that of real images in some cases, whereas incomplete or localized staining was noted in others. Moreover, a critical limitation shared by both methods was the introduction of structural distortions during style transfer, with varying degrees of alteration to the original tissue architecture. The staining results produced by ILVR and SD were entirely unacceptable. The staining was overly faint with poor color expressiveness, and more critically, severe structural distortions were introduced, rendering these results unsuitable for medical imaging applications where structural fidelity is a fundamental requirement. The virtual images generated by our model exhibited substantially superior staining quality compared with the competing algorithms, with color expressiveness more closely approximating that of real ICC images. Importantly, the structural information of the input images was fully preserved without discernible alterations, better fulfilling the practical requirements of clinical chromoendoscopy.

Discrimination between real and virtual images by eleven endoscopists

In the authenticity discrimination task, the overall sensitivity, specificity, and accuracy of the 11 endoscopists for identifying virtual ICC images were 77.1% (95%CI: 74.3%–79.7%), 79.1% (95%CI: 76.3%–81.6%), and 78.1% (95%CI: 76.2%–79.9%), respectively, indicating that most virtual images were effectively distinguished. Stratified analysis by experience level demonstrated superior discriminative performance in the senior group over the novice group. The sensitivity, specificity, and accuracy for the senior group were 82.8% (95%CI: 78.9%–86.1%), 83.3% (95%CI: 79.5%–86.5%), and 83.1% (95%CI: 80.4%–85.4%), compared with 72.4% (95%CI: 68.3%–76.1%), 75.5% (95%CI: 71.6%–79.1%), and 74.0% (95%CI: 71.2%–76.6%) for the novice group.

Endoscopists scoring results for real and virtual images

Endoscopist scores for virtual and real ICC images are presented in Table 3. For overall contrast, no significant difference was found between virtual and real images in the overall assessment (3.85 ± 0.90 vs. 3.92 ± 0.95, $P = 0.059$). Stratified analysis showed that senior endoscopists assigned lower scores to virtual than to real images (3.63 ± 0.84 vs. 3.73 ± 0.94, $P = 0.045$), whereas no significant difference was observed in the novice group (4.04 ± 0.91 vs. 4.07 ± 0.94, $P = 0.530$). For lesion boundary clarity, comparable overall scores were obtained (3.72 ± 0.96 vs. 3.76 ± 0.99, $P = 0.157$). Senior endoscopists rated virtual images significantly lower than real images (3.54 ± 0.90 vs. 3.69 ± 0.98, $P = 0.002$), while no significant difference was detected in the novice group (3.86 ± 0.98 vs. 3.82 ± 1.01, $P = 0.269$). For microstructure visualization, virtual images received significantly lower scores than real images across all assessments. Overall, virtual staining achieved performance approaching that of real staining in

overall contrast and lesion boundary clarity, with only subtle differences noted by senior endoscopists, whereas a notable gap remained in microstructure visualization.

Table 3
Endoscopists scoring results for real and virtual ICC images

Dimensions	Endoscopists	Values	Virtual	Real	*P*
Overall contrast	All	Mean ± SD	3.85 ± 0.90	3.92 ± 0.95	0.059
		Median (IQR)	4(3–5)	4(3–5)	
	Senior	Mean ± SD	3.63 ± 0.84	3.73 ± 0.94	0.045
		Median (IQR)	4(3–4)	4(3–4)	
	Novice	Mean ± SD	4.04 ± 0.91	4.07 ± 0.94	0.530
		Median (IQR)	4(4–5)	4(3–5)	
Lesion boundary clarity	All	Mean ± SD	3.72 ± 0.96	3.76 ± 0.99	0.157
		Median (IQR)	4(3–4)	4(3–5)	
	Senior	Mean ± SD	3.54 ± 0.90	3.69 ± 0.98	0.002
		Median (IQR)	4(3–4)	4(3–4)	
	Novice	Mean ± SD	3.86 ± 0.98	3.82 ± 1.01	0.269
		Median (IQR)	4(3–5)	4(3–5)	
Microstructure visualization	All	Mean ± SD	3.35 ± 0.96	3.61 ± 1.03	< 0.001
		Median (IQR)	3(3–4)	4(3–4)	
	Senior	Mean ± SD	3.09 ± 0.89	3.44 ± 0.99	< 0.001
		Median (IQR)	3(3–4)	3(3–4)	
	Novice	Mean ± SD	3.58 ± 0.96	3.76 ± 1.05	< 0.001
		Median (IQR)	4(3–4)	4(3–5)	

SD, standard deviation; IQR, interquartile range.

As shown in Table 4, moderate-to-good inter-rater agreement (intraclass correlation coefficient > 0.5) was achieved for both virtual and real images across all three dimensions among the 11 endoscopists. In the senior group, moderate agreement was observed for real images in overall contrast and lesion boundary clarity, but good agreement for microstructure visualization, while good agreement was achieved for virtual images across all dimensions. The novice group showed moderate agreement across all dimensions. Overall, higher agreement was found in the senior group than in the novice group, indicating more consistent judgment criteria among experienced endoscopists.

Table 4
Scoring consistency analysis by endoscopists.

Dimensions	Endoscopists	Images	Intraclass correlation coefficient (95%CI)
Overall contrast	All	Virtual	0.82(0.74–0.87)
		Real	0.80(0.71–0.86)
	Senior	Virtual	0.76(0.66–0.83)
		Real	0.68(0.56–0.78)
	Novice	Virtual	0.70(0.56–0.80)
		Real	0.66(0.50–0.77)
Lesion boundary clarity	All	Virtual	0.84(0.76–0.89)
		Real	0.81(0.74–0.87)
	Senior	Virtual	0.81(0.74–0.87)
		Real	0.74(0.63–0.82)
	Novice	Virtual	0.70(0.53–0.81)
		Real	0.63(0.44–0.75)
Microstructure visualization	All	Virtual	0.80(0.70–0.86)
		Real	0.82(0.74–0.88)
	Senior	Virtual	0.77(0.66–0.84)
		Real	0.77(0.67–0.84)
	Novice	Virtual	0.63(0.43–0.76)
		Real	0.61(0.42–0.73)

95%CI, 95% confidence intervals.

To investigate whether image type influenced scoring trends, the interaction between staining method and endoscopist group was analyzed (Fig. 2). No significant interaction was found for overall contrast ($P > 0.05$), indicating that scoring was unaffected by image type. For lesion boundary clarity, lower scores were assigned to virtual images only in the senior group, with no significant interaction in the overall or novice group analyses. For microstructure visualization, significantly lower scores were assigned to virtual images across all groups, consistent with Table 3.

Discussion

In this study, an AI model for virtual ICC was developed by integrating multicenter endoscopic image data with a purpose-optimized network architecture and a multi-dimensional evaluation framework, enabling reasonably high-quality translation from WLI to virtual ICC images. The results demonstrated that the virtual images generated by our model outperformed three categories of state-of-the-art algorithms across objective metrics, including color expressiveness, structural fidelity, and overall visual quality. Subjective evaluation further revealed that the virtual images were comparable to real ICC images in overall contrast and lesion boundary clarity, suggesting promising potential for clinical application.

Several previous studies have preliminarily explored AI-based virtual ICC. Fukuda A et al.[15] first used GAN to synthesize virtual ICC images from WLI of colorectal lesions and developed a YOLO V3-based detection model using 13,617 WLI, 5,509 real ICC, and 13,617 virtual images. However, no quality evaluation of virtual images was conducted, and the WLI-trained model outperformed the virtual-trained one, possibly due to increased brightness, reduced contrast, and greater noise in virtual images. Widya A R et al.[16] generated virtual ICC from gastroscopy frames using CycleGAN, but only compared reconstructed gastric models without evaluating staining quality. Suzuki S et al.[17] used 911 WLI images and 722 ICC images to generate virtual early gastric cancer images, finding lower margin delineation accuracy on virtual images than on WLI, with no direct comparison to real ICC. Lu et al.[18] reported comparable quality between virtual and real ICC (3.14 vs. 3.02) by five experts, but only 10 image pairs were tested. Sato R et al.[19] trained CycleGAN on 1,985 WLI and 896 ICC images and applied it to 40 cholangioscopy cases, with three experts agreeing on advantages of virtual ICC over WLI and narrow-band imaging for biliary lesion visualization. Compared with these studies, our work was systematically optimized in model design, data scale, and evaluation. At the model level, targeted improvements were made to balance staining quality and structural fidelity. At the data level, multicenter data from five hospitals were used to increase sample size and diversity. At the evaluation level, a multi-dimensional framework was established combining objective metrics and subjective clinician scoring, overcoming the limitations of single-method or small-sample evaluations in prior studies.

Regarding model architecture selection, our model was developed based on an improved CycleGAN, driven by the intrinsic requirements of virtual ICC. This task involves a complex image distribution mapping problem that demands precise reconstruction of color distribution and local contrast while preserving the original tissue architecture. Conventional colorization methods are inadequate for capturing the complex nonlinear features of endoscopic images, while diffusion models, despite their success in natural image generation, carry a risk of structural distortion in medical imaging contexts. Our model offers a more balanced capability in distribution fitting and detail preservation. Furthermore, targeted optimizations were implemented to account for the structural sensitivity of endoscopic images, ensuring that staining enhancement did not interfere with lesion morphology. The rationality of this design was validated by comparative experiments, in which our model achieved superior performance over multiple competing algorithms across FID, ΔCF, and SSIM, alongside improved structural fidelity in visual results.

In terms of experimental design, a multi-level validation strategy was adopted to ensure reliable and generalizable results. Data from five hospitals were used to reduce single-center bias. A temporal split of WCHSCU data assessed temporal generalizability, while institutional allocation of data from the other four hospitals assessed cross-center generalizability. This scheme closely reflects clinical deployment scenarios and enables comprehensive performance evaluation. Comparative experiments against three state-of-the-art algorithms further confirmed that our model achieved superior and more balanced performance in feature proximity, color expressiveness, and structural consistency.

For virtual staining quality evaluation, a systematic analysis was performed combining objective metrics and subjective assessment. Subjective assessment showed that the 11 endoscopists could distinguish virtual from real ICC images with an overall accuracy of 78.1%, suggesting residual color discrepancies likely due to incomplete mapping to the real staining distribution. However, no statistically significant differences were observed between virtual and real images in overall contrast or lesion boundary clarity scores, indicating near-real performance in macroscopic structure enhancement. Nonetheless, scores for microstructure visualization remained lower for virtual images, indicating room for improvement in fine-structure representation. Higher discriminative accuracy was observed in the senior group than in the novice group (83.1% vs. 74.0%), likely reflecting greater sensitivity to subtle structural features. This suggests that virtual staining may offer greater benefit for less experienced endoscopists, potentially narrowing the experience-related diagnostic gap. Consistency and interaction analyses confirmed moderate-to-good inter-rater agreement across all dimensions and image types, with comparable levels and no significant variation by image source, supporting model stability.

Despite these promising results, several limitations exist. First, although multicenter data were used, all images were from regional institutions. External validation on broader populations and multi-device platforms is needed to improve generalizability and clinical applicability. Second, this study used static images, whereas endoscopy is inherently a dynamic video process. Further investigation is needed on the temporal consistency and real-time processing capability of virtual staining on continuous video frames. Third, clinical evaluation was primarily based on subjective image quality scoring, without systematic assessment of its impact on actual diagnostic performance, such as improvements in detection rate and diagnostic accuracy for early gastric and colorectal cancers. Future work should focus on developing a real-time virtual staining system and conducting prospective studies in real-world settings to validate its value in enhancing diagnostic efficacy.

In conclusion, an AI model for virtual ICC was developed and validated in this study. The model generated virtual ICC images of reasonably high-quality, which were comparable to actual staining in overall contrast and lesion border definition, with stable performance and favorable generalizability across centers. By reducing reliance on conventional staining, streamlining endoscopic practice, and possibly increasing the yield of early GI malignancy identification, the present study offers a viable strategy that carries clear translational potential.

Abbreviations

WLI
White-light imaging
ICC
Indigo carmine chromoendoscopy
AI
Artificial intelligence
GANs
Generative adversarial networks
WCHSCU
West China Hospital, Sichuan University
WCFHSCU
West China Fourth Hospital, Sichuan University
ZFPH
Zigong Fourth People's Hospital
CPH
Cangxi People's Hospital
NCH
Nanchong Central Hospital
FID
Fréchet Inception Distance
CF
Colorfulness Score
SSIM
Structural Similarity Index
CI
Confidence intervals

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Biomedical Research Ethics Committee of West China Hospital, Sichuan University [Approval No. 2024 Review (1337)].

Consent for publication

Informed consent was waived for retrospective data.

Availability of data and materials

The datasets used in this study are available from the corresponding author upon reasonable request, subject to approval by the Biomedical Research Ethics Committee of West China Hospital, Sichuan

University. The source code used in this study is not publicly available but may be provided to qualified researchers upon reasonable request to the corresponding author.

Competing interests

Xianglei Yuan received funding from the National Natural Science Foundation of China (Grant No: U24A20341) and Noncommunicable Chronic Diseases-National Science and Technology Major Project (Grant No: 2026ZD0554101). Xianhui Zeng received funding from the National Natural Science Foundation of China (Grant No: 62406204) and Natural Science Foundation of Sichuan Province (Grant No: 2025ZNSFSC0486). Weili Jiang received funding from the Fundamental Research Funds for the Central Universities (Grant No: 2682026CX115) and Science and Technology Program of Tibet Autonomous Region (Grant No: XZ202601ZY0087). Ruide Liu, Yuan Gao, Wanhong Zhang, Xiaogang Bi, Tao Zhang, Linling Liu, Luli Hu, Congxin Liang, Yuhang Zhang, Liuxiang Chen, Qi Luo, Nuoya Zhou, Pinghua Wen, Huan Jiang, Bing Hu, and Yi Mou have no conflicts of interest or financial ties to disclose.

Funding

This study was supported by the National Natural Science Foundation of China (Grant No: U24A20341, 62406204), Noncommunicable Chronic Diseases-National Science and Technology Major Project (Grant No: 2026ZD0554101), Natural Science Foundation of Sichuan Province (Grant No: 2025ZNSFSC0486), Fundamental Research Funds for the Central Universities (Grant No: 2682026CX115) and Science and Technology Program of Tibet Autonomous Region (Grant No: XZ202601ZY0087).

Authors' contributions

Xianglei Yuan and Xianhui Zeng designed the study, collected, and analyzed the data, and drafted the original manuscript. Wanhong Zhang, Xiaogang Bi and Tao Zhang contributed to collected the data. Ruide Liu, Linling Liu, Luli Hu, Yuhang Zhang, Liuxiang Chen, Qi Luo, Nuoya Zhou, Pinghua Wen and Huan Jiang contributed to designing the study. Yuan Gao and Bing Hu contributed to revising the manuscript. Congxin Liang and Weili Jiang contributed to developing the artificial intelligence model. Yi Mou was the principal investigator and revised the manuscript. All authors have reviewed and approved the final version of the manuscript and have consented to its publication.

Acknowledgments

We acknowledge the funding support from the National Natural Science Foundation of China (Grant No: U24A20341, 62406204), Noncommunicable Chronic Diseases-National Science and Technology Major Project (Grant No: 2026ZD0554101), Natural Science Foundation of Sichuan Province (Grant No: 2025ZNSFSC0486), Fundamental Research Funds for the Central Universities (Grant No: 2682026CX115) and Science and Technology Program of Tibet Autonomous Region (Grant No: XZ202601ZY0087).

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63.
2. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent. 2024;4:47–53.
3. Banks M, Graham D, Jansen M, et al. British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma. Gut. 2019;68:1545–75.
4. Dai J, Shen YF, Sano Y, et al. Evaluation of narrow-band imaging in the diagnosis of colorectal lesions: is a learning curve involved? Dig Endosc. 2013;25:180–8.
5. Pohl J, Schneider A, Vogell H, et al. Pancolonic chromoendoscopy with indigo carmine versus standard colonoscopy for detection of neoplastic lesions: a randomised two-centre trial. Gut. 2011;60:485–90.
6. Matsuda T, Ono A, Sekiguchi M, et al. Advances in image enhancement in colonoscopy for detection of adenomas. Nat Rev Gastroenterol Hepatol. 2017;14:305–14.
7. Liu W, Wu Y, Yuan X, et al. Artificial intelligence-based assessments of colonoscopic withdrawal technique: a new method for measuring and enhancing the quality of fold examination. Endoscopy. 2022;54:972–9.
8. Yuan XL, Zeng XH, Liu W, et al. Artificial intelligence for detecting and delineating the extent of superficial esophageal squamous cell carcinoma and precancerous lesions under narrow-band imaging (with video). Gastrointest Endosc. 2023;97:664–72.
9. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut. 2019;68:2161–9.
10. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. Lancet Gastroenterol Hepatol. 2020;5:352–61.
11. Brophy E, Wang Z, She Q, et al. Generative adversarial networks in time series: A systematic literature review. ACM-CSUR. 2023;55(10):1–31.
12. Liu K, Li B, Wu W et al. VSGD-Net: Virtual Staining Guided Melanocyte Detection on Histopathological Images[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 1918–1927.
13. Wölflein G, Um IH, Harrison DJ et al. HoechstGAN: Virtual Lymphocyte Staining Using Generative Adversarial Networks[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 4997–5007.
14. Lyu J, Fu Y, Yang M, et al. Generative adversarial network–based noncontrast CT angiography for aorta and carotid arteries. Radiology. 2023;309(2):e230681.
15. Fukuda A, Miyamoto T, Kamba S, et al. Generating virtual chromoendoscopic images and improving detectability and classification performance of endoscopic lesions[C]//MICCAI Workshop on

Domain Adaptation and Representation Transfer. Cham: Springer International Publishing; 2019. pp. 99–107.

16. Widya AR, Monno Y, Okutomi M, et al. Stomach 3D reconstruction using virtual chromoendoscopic images[J]. IEEE J Translational Eng Health Med. 2021;9:1–11.
17. Suzuki S, Monno Y, Arai R, et al. Diagnostic performance of deep-learning-based virtual chromoendoscopy in gastric neoplasms[J]. Gastric Cancer. 2024;27(3):539–47.
18. Lu YB, Lu SC, Li FD, et al. Artificial intelligence-aided diagnostic imaging: A state-of-the-art technique in precancerous screening. J Gastroenterol Hepatol. 2024;39(3):544–51.
19. Sato R, Matsumoto K, Kinugasa H, et al. Virtual indigo carmine chromoendoscopy images: a novel modality for peroral cholangioscopy using artificial intelligence technology (with video)[J]. Gastrointest Endosc. 2024;100(5):938–46.
20. Kang X, Yang T, Ouyang W et al. Ddcolor: Towards photo-realistic image colorization via dual decoders[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2023: 328–338.
21. Zhu JY, Park T, Isola P et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. Proceedings of the IEEE international conference on computer vision, 2017: 2223–2232.
22. Ye JC, Kim B. KWON G. Unpaired image-to-image translation via neural schrödinger bridge[C]//The Twelfth International Conference on Learning Representations, ICLR 2024. International Conference on Learning Representations, 2024.
23. Choi J, Kim S, Jeong Y et al. Ilvr: Conditioning method for denoising diffusion probabilistic models[J]. arXiv preprint arXiv:2108.02938, 2021.
24. Rombach R, Blattmann A, Lorenz D et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684–10695.

Figures

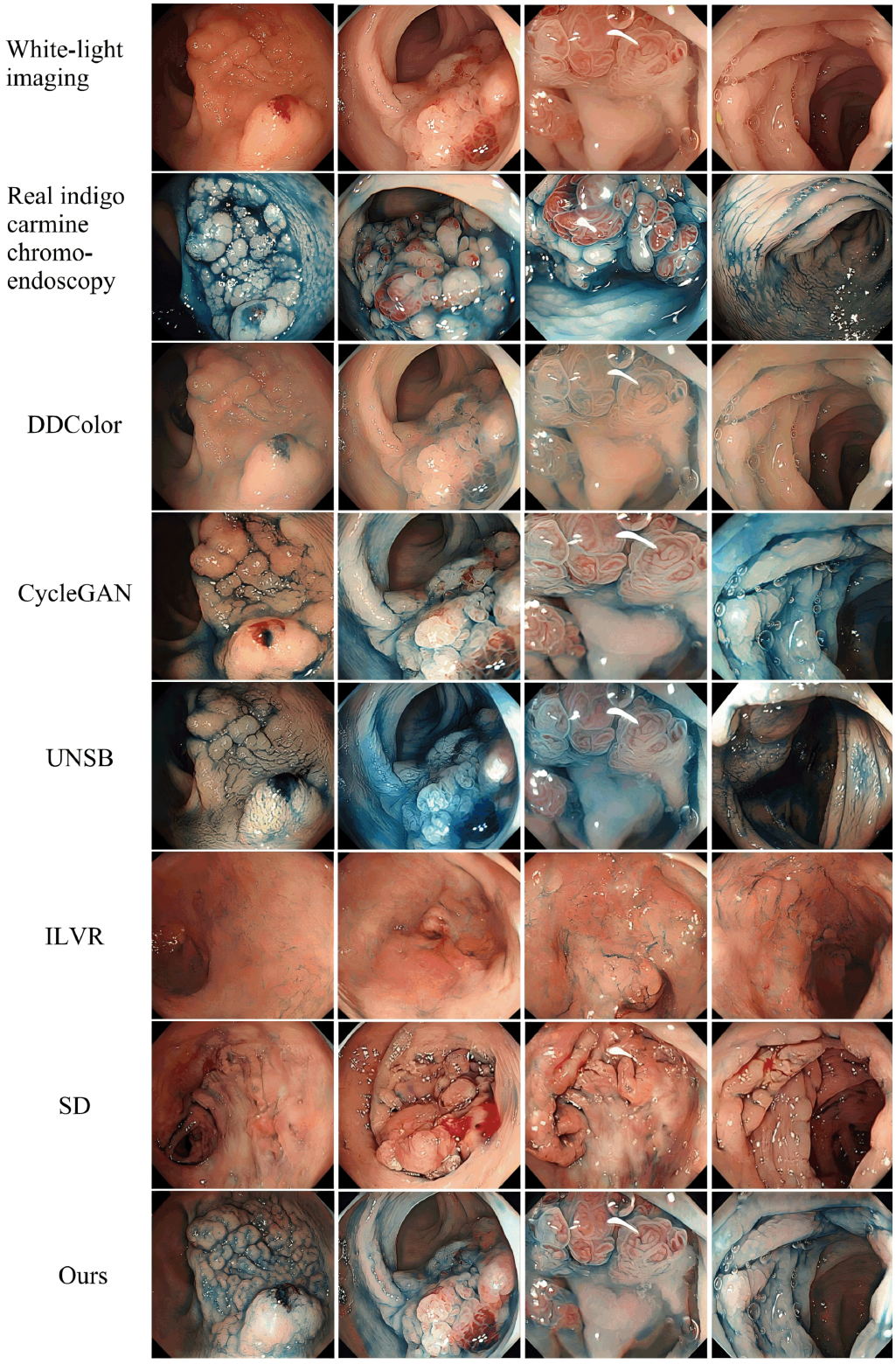


Figure 1

Examples of virtual images generated by different algorithms.

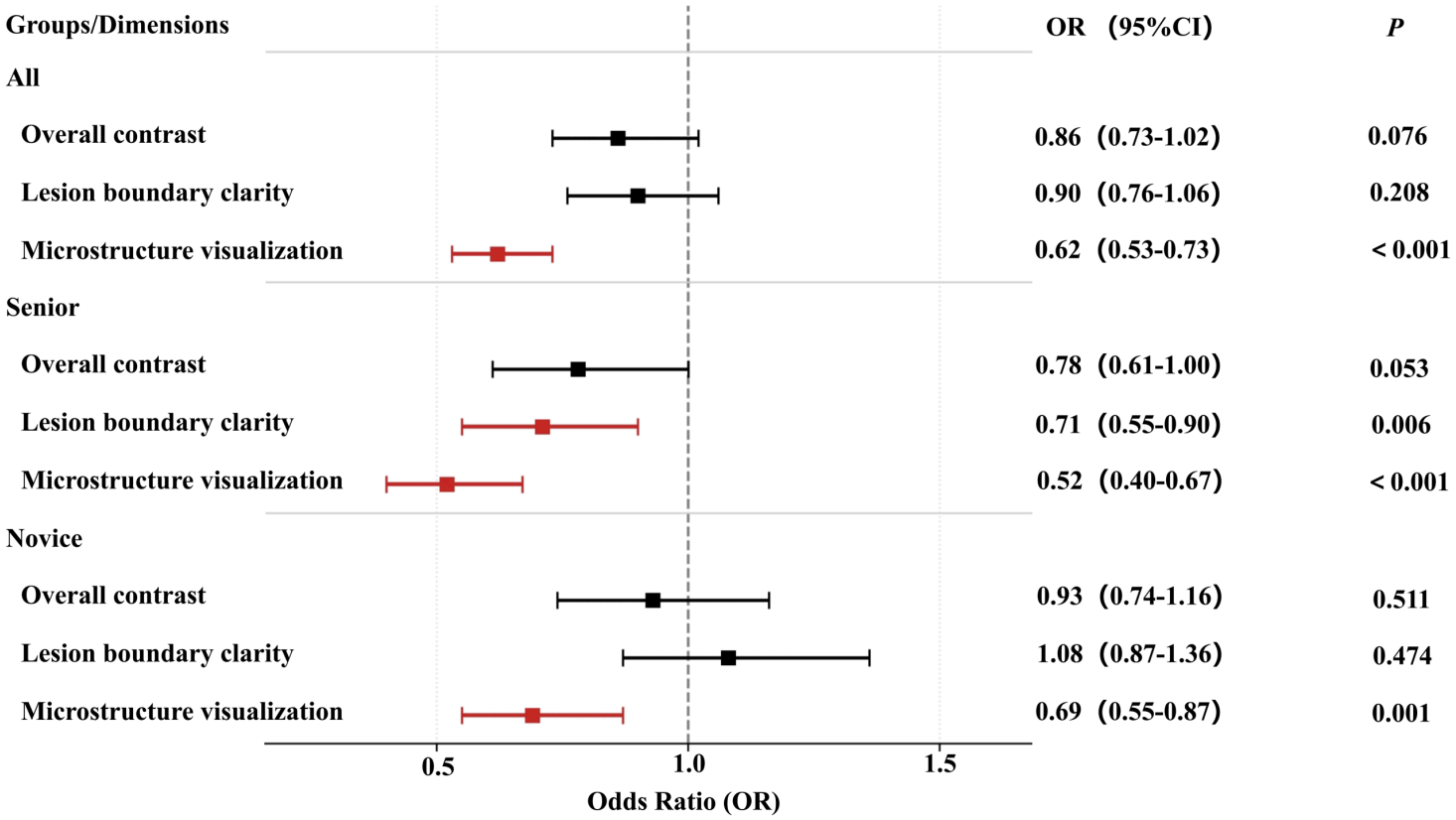


Figure 2

Forest plot of the interaction effects between virtual staining and endoscopist groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx