Journal of Pediatric Surgery 61 (2026) 163471

Contents lists available at ScienceDirect

Journal of Pediatric Surgery

journal homepage: www.sciencedirect.com/journal/journal-of-pediatric-surgery

Clinical Research Paper

Generative AI for Hirschsprung disease: Can synthetic fluorescence confocal microscopy images enhance intraoperative detection of ganglionic bowel?

Donatella Di Fabrizio [a, c, *], Agnese Sbrollini [b], Edoardo Bindi [a, c], Gaia Goteri [d], Laura Burattini [b], Giovanni Cobellis [a, c]

[a] Pediatric Surgery Unit, Salesi Children's Hospital, Polytechnic University of Marche, Ancona, Italy
[b] Department of Information Engineering, Polytechnic University of Marche, Ancona, Italy
[c] Department of Specialized Clinical and Odontostomatological Sciences, Polytechnic University of Marche, Ancona, Italy
[d] Department of Biomedical Sciences and Public Health, Institute of Pathological Anatomy, Polytechnic University of Marche, Ancona, 60126, Italy

ARTICLE INFO

Article history:
Received 11 April 2026
Received in revised form
6 July 2026
Accepted 8 September 2026

Keywords:
Hirschsprung disease
Fluorescence confocal microscopy
Generative adversarial network
Synthetic images
Artificial intelligence

ABSTRACT

Background: Accurate identification of ganglionated bowel is essential during laparoscopic pull-through for Hirschsprung Disease (HD), yet intraoperative biopsy interpretation is time-sensitive and operator-dependent. Fluorescence confocal microscopy (FCM) provides rapid imaging of fresh tissue, and deep-learning (DL) has the potential to extract diagnostic patterns from these images automatically. However, DL development is limited by HD rarity and images scarcity. Generative-AI may address this gap by synthesizing realistic images to expand underrepresented data, improving model robustness. This study aims to evaluate whether conditional generative adversarial network (cGAN)—synthetic FCM images can improve DL performance for discriminating ganglionic cells.

Methods: Between November 2024 and November 2025, intraoperative FCM was routinely performed during laparoscopic pull-through for HD. High-resolution images were annotated to identify ganglionic and aganglionic regions, subdivided into 256 × 256-pixel tiles (resolution 0.5 × 0.5 μm/pixel), and grayscale (128 levels) converted. Three convolutional neural networks (CNN), having the same architecture, were trained on different datasets: CNN-1 considered only real tiles (461 ganglionic and 1374 aganglionic), CNN-2 considered the same real tiles (461 ganglionic and 1374 aganglionic) and 913 conditional-GAN simulated ganglionic tiles, and CNN-3 considered the same real tiles (461 ganglionic and 1374 aganglionic) and 913 data augmented ganglionic tiles. CNNs were validated in the same independent testing dataset composed of real 590 aganglionic and 198 ganglionic tiles.

Results: CNN-1, CNN-2 and CNN-3 identified ganglionic tiles with 75%, 84% and 80% accuracy, 78%, 91% and 79% sensitivity, 73%, 78% and 75% specificity, respectively. CNN-2 achieved a significantly higher ROC-AUC than CNN-1 (0.94 vs 0.84 and 0.80; $P < 0.01$).

Conclusion: Synthetic augmentation with cGAN-generated ganglionic tiles significantly improved CNN discrimination of ganglionic bowel on real FCM images, supporting generative-AI as a scalable strategy to strengthen DL for discriminating ganglionic bowel in HD.

© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Hirschsprung disease (HD) remains one of the most challenging disorders of the enteric nervous system, because of its rarity, heterogeneity, and histologically subtleness [1,2]. The diagnosis hinges on identifying the absence of ganglion cells and the presence of hypertrophic nerve trunks across the myenteric and submucosal plexuses, and even in expert hands, the conventional histopathological workflow is slow and vulnerable to sampling bias, because millions of tissue structures must be inspected in each case, and interobserver variability remains high [3,4]. These constraints become even more pronounced in the intraoperative setting, where the surgeon depends on rapid evaluation of sequential biopsies to define the pull-through level.

Recent progress in digital pathology (DP) and artificial intelligence (AI) has begun to erode these limitations. Multiple groups

* Corresponding author.
E-mail address: donatella.difabrizio@ospedaliriuniti.marche.it (D. Di Fabrizio).

https://doi.org/10.1016/j.jpedsurg.2026.163471
0022-3468/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

have demonstrated that machine-learning (ML) models can support or even outperform human experts in detecting ganglion cells, segmenting neural structures, and reducing diagnostic time by orders of magnitude. Greenberg et al. showed that an AI system trained on 95 cases could detect ganglion cells with 96% sensitivity and 99% specificity, while reducing pathologist workload by >95%, despite the rarity of training data and the overwhelming imbalance between relevant and irrelevant candidate pixels [5]. More recently, Duci et al. developed a U-Net-based system achieving >91% accuracy in differentiating ganglia, normal nerves, and hypertrophic nerves in HD pull-through specimens, highlighting the feasibility of deep learning even in small pediatric datasets [6]. Despite rising expectations for AI in pediatric and neonatal care, the limited, heterogeneous, and hard-to-standardize datasets available in this field rarely allow external validation and often cause a 15–20% drop in performance across institutions [7]. HD pathology therefore exemplifies the broader pediatric AI paradox: the conditions that would benefit most from automation are those with the least data available to enable it.

Generative adversarial networks (GANs) offer a theoretical and practical breakthrough for this tension. Instead of relying solely on the limited number of real histological samples, GANs learn the underlying distribution of image features and can produce new, high-fidelity synthetic images that expand the data manifold without compromising patient privacy [8]. When applied to ex-vivo fluorescence confocal microscopy (FCM), capable of visualizing ganglion cells and hypertrophic nerve fibers in fresh tissue within minutes [9], GAN-generated images could dramatically amplify the effective training set available for downstream classifiers and reduce overfitting [8].

This study explores whether GAN-synthetic high-resolution FCM images can effectively boost the limited real dataset and strengthen deep-learning performance for discriminating ganglionic cells in HD.

2. Methods

2.1. Patient cohort

From November 2024 to November 2025, intraoperative FCM was routinely performed during pull-through procedures for HD at our center. This imaging step is integrated into the surgical pathway and provides rapid microscopic visualization of fresh bowel without altering the tissue or interfering with its subsequent definitive pathological assessment.

For this study, we retrospectively selected anonymized FCM high resolution digital imaging obtained during standard care. Scans were excluded when markedly affected by cautery, mechanical distortion, or inadequate fluorescence staining. Basic clinical variables, such as patient age, indication for biopsy, and final histological classification (ganglionic or aganglionic), were recorded solely to contextualize the imaging dataset. No additional sampling or interventions were introduced beyond standard surgical practice.

2.2. Sample handling and FCM imaging

Immediately after removal, each biopsy was gently rinsed in saline to clear surface blood, briefly dipped in ethanol, and sequentially stained with acridine orange (0.04%) and fast green, followed by a final saline rinse to remove excess dye. The tissue was then imaged using the VivaScope 2500 Multilaser platform (Caliber I&D, Rochester, NY, USA), which generates high-resolution mosaics by optically sectioning fresh, unfixed tissue and stitching adjacent fields into a continuous large-area map. Acquisition time varied from 30 to 90 s depending on specimen dimensions.

The resulting mosaics were archived in their native format and later used for digital annotations and computational processing. Dataset splitting was performed at tissue-sample level. Tiles derived from the same tissue sample were not shared between training and testing datasets, in order to reduce the risk of data leakage due to visually correlated tiles originating from the same specimen.

All tissue specimens subsequently underwent routine histological examination without deviation from normal workflow.

2.3. Annotation strategy and tile extraction

The FCM mosaics were imported into QuPath (v6.0), where a pediatric surgeon experienced in HD, together with a pediatric expert pathologist, reviewed each scan. Regions displaying clear evidence of ganglionated bowel were outlined separately from areas showing hypertrophic nerves and absence of ganglion cells. Fields rendered uninterpretable by artifacts or excessive background noise were excluded. Annotations were iteratively reviewed until full agreement was reached.

Following annotation, each mosaic was divided into small square tiles to create a structured dataset suitable for deep-learning workflows. Tiles of 256 × 256 pixels were exported at a uniform spatial resolution of 0.5 μm/pixel. Only tiles overlapping an annotated region were retained; empty background areas were discarded. Because multiple tiles were extracted from each tissue sample, tile counts were not considered independent biological observations. Tile-level classification was therefore used as a technical endpoint to compare model behavior under different training strategies.

2.4. Image simulation by a conditional generative adversarial network

The used generative adversarial network is a conditional one (cGAN). The cGAN is a type of GAN that allows the addition of extra inputs, such as class labels (i.e., ganglionic or aganglionic), to both the generator and discriminator as a condition to guide the data generation. Both networks are conditioned by auxiliary information with different contextual conditions, and therefore, in the ideal cGAN model, the generator learns to generate data with specific conditions rather than random input/noise, and the discriminator learns to discern the authenticity of the input based on the provided condition.

In our case, the generator is a Mixture-of-Experts (MoE) network composed of three experts' sub-networks. Each expert is a convolutional neural network (CNN) composed of residual blocks (two 3 × 3 convolutions with BatchNorm and ReLU) connected by self-Attention Layers. The outputs of the experts are combined with a learned gate network, which outputs a softmax over the experts, dynamically weighing their contributions based on the input. Combined with the generator, the discriminator is a PatchGAN-style CNN with 4 convolutional layers, BatchNorm, and LeakyReLU.

The training of the generator (L_G) leverages a combination of four complementary losses, each targeting a different aspect of image quality. The first is an adversarial loss (L_{adv}), which encourages the generator to produce images indistinguishable from real images. The second is the classification loss (L_{cls}), that based on cross-entropy, enforces conditional generation aligned with the input class label. The third one is the histogram loss (L_{hist}), which aim to preserve the color distribution in the generated images, trying to force the generated image's intensity distribution to match that of real images. Finally, the fourth is the perceptual loss (L_{perc}), which ensures similarity in high-level features between

generated and real images, preserving the global structure and semantic content. Thus, the total generator loss is:

$$L_G = \lambda_{adv}L_{adv} + \lambda_{cls}L_{cls} + \lambda_{hist}L_{hist} + \lambda_{perc}L_{perc}$$

with empirically chosen weights $\lambda_{adv} = 1.0$, $\lambda_{cls} = 1.0$, $\lambda_{hist} = 5.0$, and $\lambda_{perc} = 0.1$.

The discriminator loss of the discriminator consists of the sum of the adversarial and the classification losses on its output. Training was performed for 100 epochs using Adam optimizer with learning rates $1e^{-4}$ for the generator and $2e^{-4}$ for the discriminator. A batch size of 16 allowed stable training of the Mix-of-Expert generator with self-attention and perceptual loss.

2.5. Assessment of the role of generated images in classification

To assess the role of generated images in the HD ganglionic image identification, three convolutional neural networks (CNN-1 for the first approach; CNN-2 for the second approach; CNN-3 for the third approach), having the same architecture, were trained on different datasets. The methodological approaches are graphically described in Fig. 1.

The architecture of the used CNN consists of three convolutional blocks designed to progressively extract hierarchical spatial features from input images. The first block applies a 2D convolutional layer with 32 filters and a 3×3 kernel using the ReLU activation function, followed by a 2×2 max-pooling layer to reduce spatial dimensionality and computational complexity. The second and third blocks follow the same structure, with 64 and 128 convolutional filters, respectively, allowing the network to capture increasingly complex visual patterns. After feature extraction, the resulting feature maps are flattened and considered as input for a fully connected dense layer with 128 neurons and ReLU activation. To preserve overfitting, a dropout layer with a rate of 0.5 was applied during training.

The difference between the different approaches are the images as input to train the three CNN. Specifically, CNN-1 considered only real tiles, CNN-2 considered the same real tiles and the conditional-GAN simulated ganglionic tiles, and CNN-3 considered the same real tiles and augmented ganglionic tiles generated by standard data augmentation (real images using random horizontal/vertical flipping and rotation transformations). Both CNNs were tested in the same independent testing dataset.

2.6. Statistical analysis

Performance of cGANN in simulating the ganglionic images are evaluated in terms of Structural Similarity Index (SSIM), defined as a perceptual metric that quantifies image quality by measuring the similarity between two images in terms of luminance, contrast,

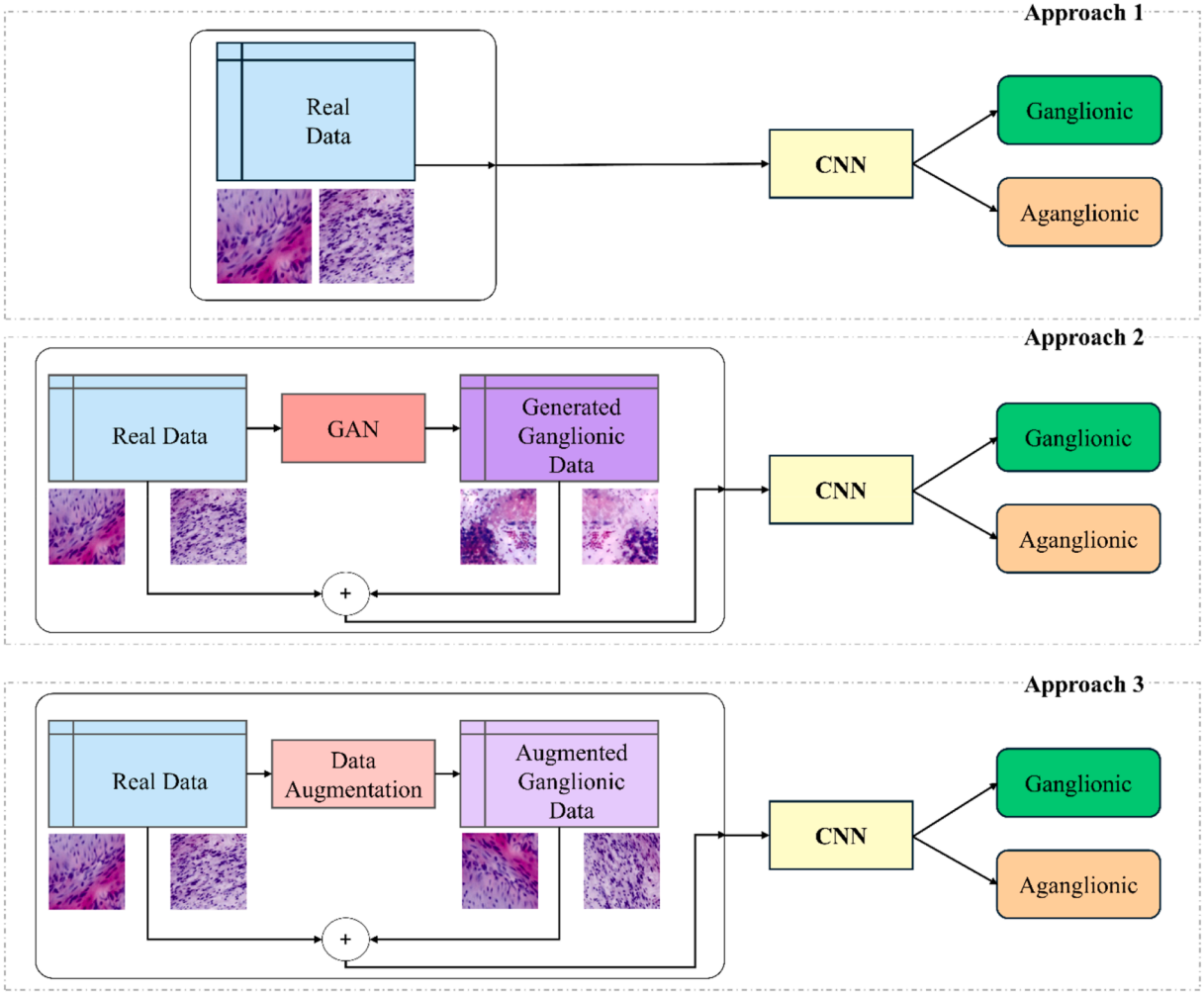

Fig. 1. Methodological approaches to assess the role of generating images on ganglionic image identification.

and structural information. Performance of the three CNN were assessed in terms of accuracy, sensitivity and specificity. Moreover, the receiver operating characteristics (ROC) are computed by quantifying its performance in terms of area under the curve (AUC). ROC of the three CNN are compared by the DeLong's test. Additionally, precision–recall curves were computed. Statistical significance was set at 0.05.

3. Results

A total of 35 tissue samples were included in the final real-image dataset. According to an approximately 70/30 split, 24 tissue samples were allocated to the training set and 11 tissue samples to the independent test set. The training set included 6 ganglionic and 18 aganglionic tissue samples, from which 1835 real FCM tiles were extracted, comprising 461 ganglionic and 1374 aganglionic tiles. The independent test set included 3 ganglionic and 8 aganglionic tissue samples, from which 788 real FCM tiles were extracted, comprising 198 ganglionic and 590 aganglionic tiles. No tissue sample contributed tiles to both the training and test datasets. CNN-1 was trained using only the 1835 real training tiles, whereas CNN-2 was trained on the same real dataset with the addition of 913 cGAN-generated ganglionic tiles, resulting in 2748 training tiles. Both models were evaluated exclusively on real tiles from the independent test set. The composition of the training and test datasets is detailed in Table 1.

The cGAN achieved moderate structural similarity in image generation, with an SSIM of 0.52 for ganglionic images and 0.54 for aganglionic images. On the independent real test set (Table 2), CNN-1 reached 75% accuracy, 78% sensitivity, and 73% specificity for ganglionic tile identification, CNN-2 achieved 84% accuracy, 91% sensitivity, and 78% specificity, and CNN-3 achieved 80% accuracy, 79% sensitivity, and 75% specificity. ROC and precision–recall curves for both models are shown in Fig. 2 and Fig. 3, respectively. CNN-2 achieved a significantly higher ROC-AUC than CNN-1 (0.94 vs 0.84, $P < 0.01$) and than CNN-2 (0.94 vs 0.80, $P < 0.01$).

Table 2
Metrics performance for the two convolutional neural networks (CNN-1, CNN-2 and CNN-3) on the three datasets.

		Training	Validation	Testing
CNN-1	**Accuracy**	80%	86%	75%
	Sensitivity	85%	73%	78%
	Specificity	74%	98%	73%
CNN-2	**Accuracy**	100%	88%	84%
	Sensitivity	99%	86%	91%
	Specificity	100%	90%	78%
CNN-3	**Accuracy**	98%	85%	80%
	Sensitivity	100%	72%	79%
	Specificity	99%	96%	75%

4. Discussion

This study shows that synthetic FCM images can improve deep-learning performance for intraoperative HD assessment in a setting where the main constraint is limited data availability. Using the same architecture and the same independent real-world test set, the model trained with real images plus cGAN-generated ganglionic tiles significantly outperformed the model trained on real images alone, with improvements in accuracy, sensitivity, specificity, and ROC-AUC. The most relevant gain was observed in sensitivity, which increased from 78% to 91%, suggesting that synthetic augmentation primarily strengthened recognition of the minority and clinically decisive class rather than simply increasing overall model confidence. An additional experiment (approach 3) using standard augmentation of ganglionic tiles achieved lower performance than the cGAN-augmented model, suggesting that the observed benefit cannot be explained solely by increased class representation. This finding supports the hypothesis that cGAN-generated images may provide additional morphological variability beyond that obtained through simple geometric transformations.

This finding is particularly meaningful in HD, where AI encounters a rare disease with sparse diagnostically informative structures within a specimen, and time-consuming expert annotation [10]. Previous studies have already shown that AI can support Hirschsprung diagnosis on conventional histology. Greenberg et al. [5] reported high sensitivity and specificity for automated ganglion cell detection, while Duci et al. [6] showed that deep learning can identify ganglia and nerve patterns with accuracies above 91% even in relatively small pediatric datasets. Braun et al. [11] further demonstrated that AI analysis of acetylcholinesterase staining can achieve excellent diagnostic performance, and Demir et al. [12] recently showed that AI-assisted detection of ganglion cells on frozen sections can improve both diagnostic accuracy and reporting time in the intraoperative setting. At the same time, the first scoping review dedicated to AI in Hirschsprung disease concluded that the field remains early in development, with few studies, small datasets, methodological heterogeneity, and limited external validation [13]. Our study fits within this emerging landscape but adds a distinct contribution: instead of only learning from scarce data, it uses generative AI to expand a diagnostically critical class that is intrinsically underrepresented.

A second major element of novelty is the use of FCM itself [14–16], increasingly being explored as a practical tool for rapid intraoperative assessment because it allows imaging of fresh, unfixed tissue without compromising subsequent standard histopathology [9]. This makes it particularly attractive for Hirschsprung disease, where surgical decision-making depends on timely identification of ganglionated bowel. A recent feasibility study showed that FCM can distinguish normal and affected bowel with diagnostic agreement comparable to frozen section in assessable samples, supporting its role as a rapid intraoperative adjunct [17]. The present work extends the field one step further, indeed the first study applying GAN-based synthetic augmentation to intraoperative FCM Hirschsprung imaging, and therefore

Table 1
Data division for the training and the testing of the two convolutional neural networks (CNN-1, CNN-2 and CNN-3).

	Training			Testing
	Approach 1	Approach 2	Approach 3	
Ganglionic	461 real images	1374 images: • 461 real images • 913 simulated images by cGANN	1374 images: • 461 real images 913 simulated images by data augmentation	198 real images
Aganglionic	1374 real images	1374 real images	1374 real images	590 real images
Overall	1835 images	2748 images	2748 images	788 images

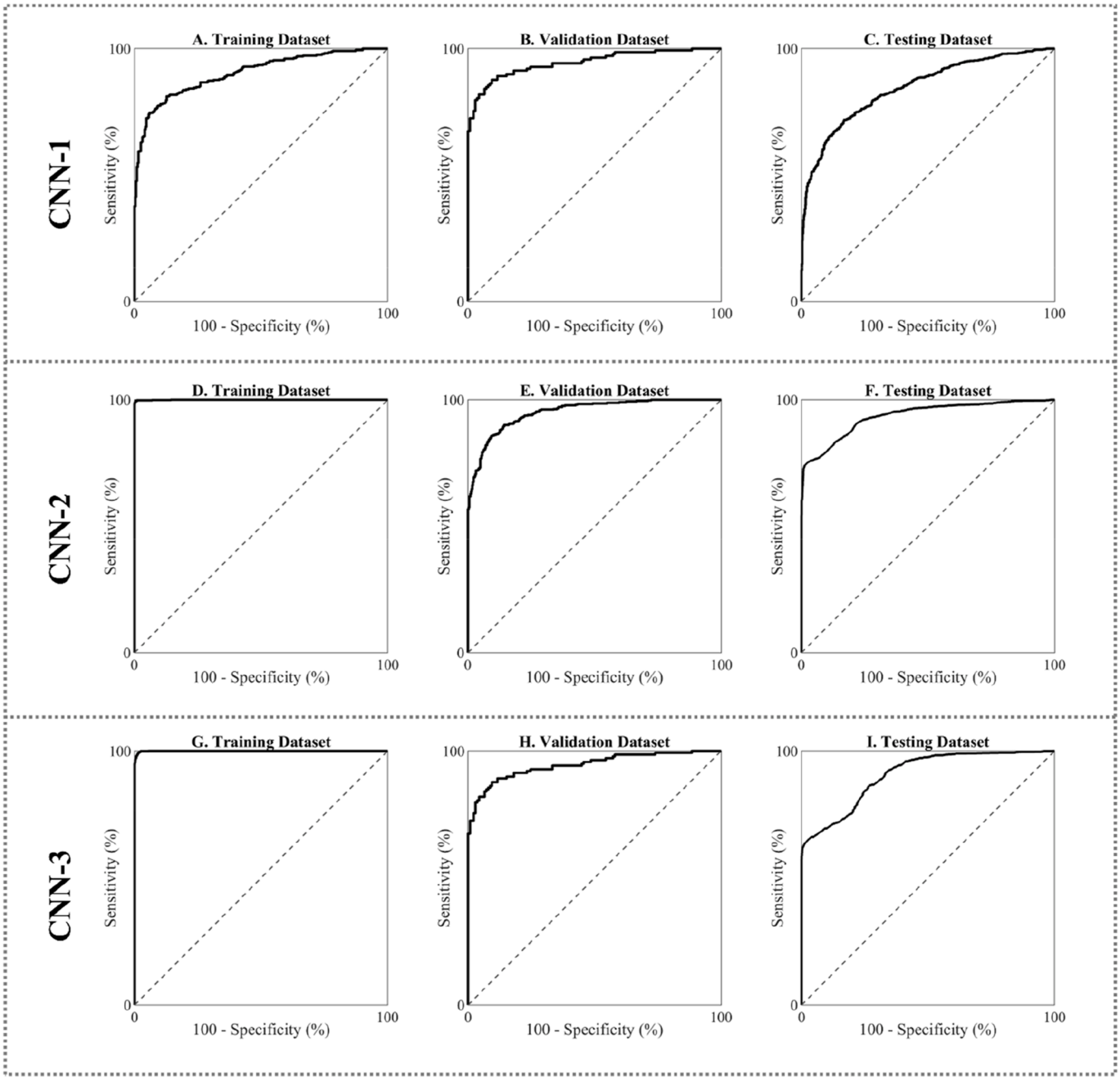

Fig. 2. Receiver operating characteristic (ROC) curves of the three convolutional neural networks (CNN).

the first attempt to address computationally the scarcity of FCM data in this specific setting.

Reviews of GANs in digital histopathology consistently suggest that their most credible role is not to replace real pathology images, but to support focused, high-value tasks such as augmentation of rare, imbalanced, or difficult-to-annotate datasets [8,18,19]. Indeed in our study, by enriching the ganglionated bowel underrepresented images with synthetic samples, the model appears to have learned a broader representation of ganglionic morphology and improved its ability to recognize it in real unseen FCM tiles, thereby making a rare morphology more learnable. In this proof-of-concept study, the moderate structural similarity of the generated images should therefore be interpreted in light of the study aim, which was not to produce fully realistic de novo images, but to generate samples capable of enriching training and improving downstream performance on real-world data. Taken together, these findings suggest that targeted synthetic augmentation may strengthen AI models built on FCM despite the very small datasets typical of pediatric surgery.

This study has several limitations. First, it is a single-center study based on a relatively small dataset acquired with one imaging platform and one acquisition workflow, which limits generalizability, and although training and testing datasets were separated at tissue-sample level, the present study did not assess

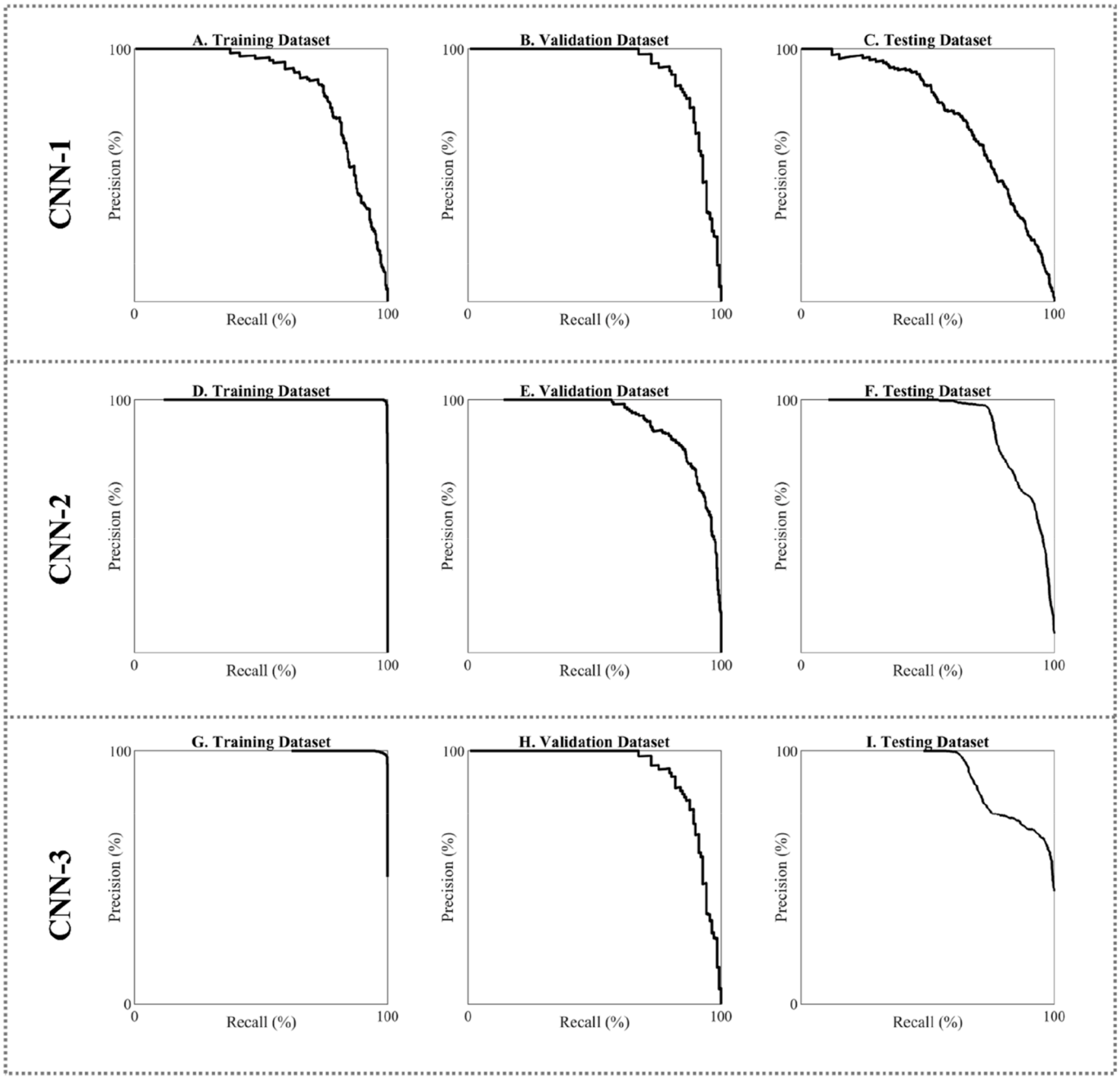

Fig. 3. Precision-recall curves of the three convolutional neural networks (CNN).

patient-level performance. Second, although the classifier was evaluated on an independent real dataset, validation remained internal from an institutional perspective. Third, the generative component was assessed mainly through its downstream contribution to classification performance rather than through a comprehensive panel of dedicated generative-validation methods. Another limitation is the absence of a formal explainability analysis. Future studies should assess the possibility of integrating explainability methods (such as Grad-CAM) to better characterize the biological basis of model decisions. Finally, as highlighted by the current Hirschsprung AI literature, the field as a whole is still limited by small datasets, heterogeneity, and limited explainability, and our study shares part of that immaturity [13,20]. For these reasons, the present work should be interpreted as a proof of concept with translational potential rather than as a deployment-ready diagnostic system.

5. Conclusion

Conditional generative adversarial network generated ganglionic fluorescence confocal microscopy images improved deep-learning discrimination of ganglionated bowel on real intraoperative images in Hirschsprung disease. This is relevant not only because the gain was achieved in a rare and data-limited setting, but also because it was built on an imaging modality that is itself emerging as a rapid intraoperative diagnostic tool. The main message is that in Hirschsprung disease, generative artificial intelligence may be most useful not when it tries to imitate

pathology for its own sake, but when it helps scarce intraoperative pathology become learnable.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

Authors report no conflict of interest.

References

[1] Montalva L, Cheng LS, Kapur R, Langer JC, Berrebi D, Kyrklund K, et al. Hirschsprung disease. Nat Rev Dis Primers 2023 Oct 12;9(1):54. https://doi.org/10.1038/s41572-023-00465-y. PMID: 37828049.

[2] Tam PKH, Chung PHY, St Peter SD, Gayer CP, Ford HR, Tam GCH, et al. Advances in paediatric gastroenterology. Lancet 2017 Sep 9;390(10099):1072–82. https://doi.org/10.1016/S0140-6736(17)32284-5. PMID: 2890 1937.

[3] Matsukuma K, Gui D, Saadai P. Hirschsprung disease for the practicing surgical pathologist. Am J Clin Pathol 2023 Mar 13;159(3):228–41. https://doi.org/10.1093/ajcp/aqac141. PMID: 36565211.

[4] Claxton HL, Lounis SA, Stanton M, Hall NJ, Aldeiri B. The diagnostic value of immunohistochemistry markers in Hirschsprung disease; A systematic review and meta-analysis. J Pediatr Surg 2025 Feb;60(2):162010. https://doi.org/10.1016/j.jpedsurg.2024.162010. Epub 2024 Oct 17. PMID: 39461141.

[5] Greenberg A, Aizic A, Zubkov A, Borsekofsky S, Hagege RR, Hershkovitz D. Automatic ganglion cell detection for improving the efficiency and accuracy of hirschprung disease diagnosis. Sci Rep 2021 Feb 8;11(1):3306. https://doi.org/10.1038/s41598-021-82869-y. PMID: 33558593; PMCID: PMC7870950.

[6] Duci M, Magoni A, Santoro L, Dei Tos AP, Gamba P, Uccheddu F, et al. Enhancing diagnosis of Hirschsprung's disease using deep learning from histological sections of post pull-through specimens: preliminary results. Pediatr Surg Int 2023 Nov 29;40(1):12. https://doi.org/10.1007/s00383-023-05590-z. PMID: 38019366; PMCID: PMC10687181.

[7] Wang Q, Yin J, Zhang X, Ou H, Li F, Zhang Y, et al. Applications of artificial intelligence in early childhood health management: a systematic review from fetal to pediatric periods. Front Pediatr 2025 Sep 16;13:1613150. https://doi.org/10.3389/fped.2025.1613150. PMID: 41036208; PMCID: PMC12479323.

[8] Alajaji SA, Khoury ZH, Elgharib M, Saeed M, Ahmed ARH, Khan MB, et al. Generative adversarial networks in digital histopathology: current applications, limitations, ethical considerations, and future directions. Mod Pathol 2024 Jan;37(1):100369. https://doi.org/10.1016/j.modpat.2023.100369. Epub 2023 Oct 27. PMID: 37890670.

[9] Di Fabrizio D, Tavolario I, Mastroberti F, Bindi E, Filosa A, Goteri G, et al. Use of intraoperative Ex vivo fluorescence confocal microscopy to identify ganglionic bowel in Hirschsprung's disease: a case series. J Pediatr Surg Case Rep 2025;120:103064.

[10] Schilling F, Geppert CE, Strehl J, Hartmann A, Kuerten S, Brehmer A, et al. Digital pathology imaging and computer-aided diagnostics as a novel tool for standardization of evaluation of aganglionic megacolon (Hirschsprung disease) histopathology. Cell Tissue Res 2019 Feb;375(2):371–81. https://doi.org/10.1007/s00441-018-2911-1. Epub 2018 Sep 3. PMID: 30175382.

[11] Braun Y, Friedmacher F, Theilen TM, Fiegel HC, Weber K, Harter PN, et al. Diagnosis of Hirschsprung disease by analyzing acetylcholinesterase staining using artificial intelligence. J Pediatr Gastroenterol Nutr 2024 Sep;79(3):729–37. https://doi.org/10.1002/jpn3.12339. Epub 2024 Aug 9. PMID: 39118474.

[12] Demir D, Ozyoruk KB, Durusoy Y, Cinar E, Serin G, Basak K, et al. The future of surgical diagnostics: artificial intelligence-enhanced detection of ganglion cells for Hirschsprung disease. Lab Invest 2025 Feb;105(2):102189. https://doi.org/10.1016/j.labinv.2024.102189. Epub 2024 Nov 20. PMID: 39577743.

[13] Alzate-Ricaurte S, Ocampo Osorio F, Costa-Roig A, Gómez Ayala MC, Gomez IE, Arnold MA, et al. Artificial intelligence in the diagnosis of Hirschsprung disease: a scoping review and rationale for a multicentric approach. J Pediatr Surg 2026 Jan 19;61(4):162932. https://doi.org/10.1016/j.jpedsurg.2026.162932. Epub ahead of print. PMID: 41565153.

[14] Di Fabrizio D, Bindi E, Ilari M, Filosa A, Goteri G, Cobellis G. Ex vivo fluorescence confocal microscopy meets innovation and revolutionary technology, for "Real-Time" histological evaluation, in pediatric surgical oncology. Children 2024 Nov 23;11(12):1417. https://doi.org/10.3390/children11121417. PMID: 39767846; PMCID: PMC11674610.

[15] Di Fabrizio D, Mastroberti F, Tavolario I, Bindi E, Ilari M, Marinelli M, et al. Evaluating Ex vivo fluorescence confocal microscopy for intraoperative diagnosis in pediatric surgery: a feasibility study. Eur J Pediatr Surg 2026 Feb;36(1):65–71. https://doi.org/10.1055/a-2676-2999. Epub 2025 Aug 5. PMID: 40763816.

[16] Di Fabrizio D, Daziani G, Qose I, Bindi E, Ilari M, Filosa A, et al. Seeing beyond the microscope: artificial intelligence and fluorescence confocal digital imaging in pediatric surgical pathology. Children 2025 Nov 26;12(12):1608. https://doi.org/10.3390/children12121608. PMID: 41462749; PMCID: PMC12731743.

[17] Braun Y, Gradhand E, Friedmacher F, Walczak B, Fiegel HC, Theilen TM, et al. Diagnostic assessment of Hirschsprung disease using fluorescence confocal microscopy: a feasibility study. J Pediatr Surg 2026 Jan 30;61(5):162973. https://doi.org/10.1016/j.jpedsurg.2026.162973. Epub ahead of print. PMID: 41621586.

[18] Brodsky V, Ullah E, Bychkov A, Song AH, Walk EE, Louis P, et al. Generative artificial intelligence in anatomic pathology. Arch Pathol Lab Med 2025 Apr 1;149(4):298–318. https://doi.org/10.5858/arpa.2024-0215-RA. PMID: 39836377.

[19] Laguarta JR, Jiménez VY, Blanco JL, González I. Generative images of microscopy cell arrangements using AI conditional training. International conference on bioinformatics and biomedicine (BIBM). 2024. p. 1–8. https://doi.org/10.1109/BIBM62325.2024.10947611.

[20] Elsayed H, Singer G, Till T, Till H. Latest developments in artificial intelligence and machine learning models in general pediatric surgery. Eur J Pediatr Surg 2026 Apr;36(2):157–64. https://doi.org/10.1055/a-2689-8280. Epub 2025 Aug 26. PMID: 40858266; PMCID: PMC12965793.