Fully Automated CT-Based Differential Diagnosis of Bowel Wall Thickening

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Bowel wall thickening (BWT) on CT occurs in both colon cancer and diverticulitis, two conditions requiring fundamentally different treatment. Their overlapping imaging appearance makes reliable differentiation difficult, particularly in emergency settings where rapid triage is critical. Existing deep learning approaches address this problem only partially, either targeting cancer-specific segmentation or classifying pathology within manually defined regions of interest. We present the first fully automated end-to-end framework combining 3D nnU-Net lesion localization with dual-channel 3D ResNet-18 classification. The pipeline achieved a malignancy sensitivity of 0.98 and specificity of 0.81 on the internal test set. External validation supported generalization, reaching up to 0.98 accuracy on public cancer datasets and maintaining a malignancy sensitivity of 0.74 under severe distribution shift in a heterogeneous multi-scanner [ANONYMIZED] cohort, where four 3D baselines performed near chance. Input masking experiments showed radiologically plausible decision patterns, while segmentation failure analysis suggested that model–reference disagreements mainly reflect annotation incompleteness and boundary ambiguity rather than failure to characterize BWT features.

Keywords: bowel wall thickening · abdominal CT · segmentation · classification · differential diagnosis · colon cancer · diverticulitis

1 Introduction

Bowel wall thickening (BWT) is among the most frequent CT findings in patients presenting with acute abdominal symptoms. Among its diverse etiologies, colon cancer and diverticulitis are the most prevalent and clinically consequential [7]. They require fundamentally different management, yet their overlapping clinical findings and imaging appearance make reliable differentiation a persistent challenge [26,19,5].

Even with established diagnostic criteria, CT alone frequently fails to resolve this ambiguity [4]. Diagnostic failures arise when inflammatory changes or pericolic abscesses mask underlying malignancy, or when eccentric thickening mimics benign disease. Patient-specific anatomical variability, interobserver differences, and technical factors such as bowel distension and motion artifacts further compound the problem. Missing a colon cancer diagnosis delays treatment, whereas

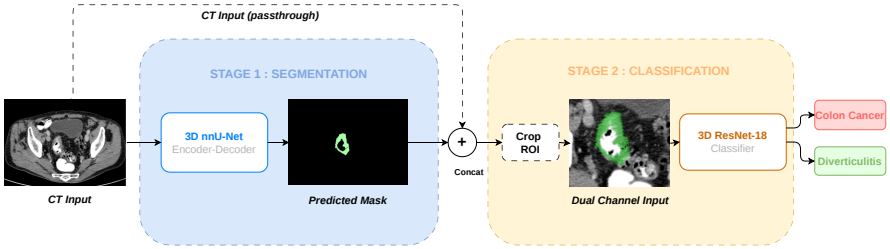


Fig. 1: Overview of the proposed two-stage framework for differential diagnosis of bowel wall thickening. A 3D nnU-Net segments regions of pathological BWT from CT volumes. The predicted segmentation mask is concatenated with the original CT as a dual-channel input and cropped to the region of interest (ROI). A 3D ResNet-18 classifier then processes the cropped volume to distinguish colon cancer from diverticulitis.

misinterpreting diverticulitis can lead to unnecessary invasive procedures. Because diverticulitis carries a non-negligible risk of occult malignancy, particularly in complicated [16] or left-sided [15] presentations, routine follow-up colonoscopy remains widely recommended despite its invasiveness [2,14,6].

Deep learning has shown promise in medical imaging, yet existing work in colorectal imaging does not address this differential diagnostic task. Published approaches either focus exclusively on tumor-specific segmentation [18,21] or classify pathology within manually defined regions of interest (ROIs) [26], limiting scalability.

To bridge this gap, we propose the first fully automated framework for CT-based differential diagnosis of BWT. The pipeline operates in two stages. A 3D nnU-Net localizes pathological BWT, and a ResNet-based classifier subsequently determines whether the underlying etiology is malignant or inflammatory.

2 Methodology

2.1 Data

Patients undergoing surgery for colon cancer or diverticulitis at a tertiary referral center (2010–2025) were retrospectively identified from a prospective database. Inclusion criteria were histopathological confirmation, venous-phase CT within 30 days preoperatively, and segmental colonic wall thickening on CT. The final cohort included 418 colon cancer cases (median age 68, 57% male) and 377 diverticulitis cases (median age 59, 55% male), split into training ($n = 718$) and test ($n = 77$) sets. Ethical approval was obtained from the institutional review board (IRB No. XXX-X-XX).

Reference BWT masks were generated via a human-in-the-loop workflow: initial annotations trained a preliminary model whose predictions were iteratively

refined by a medical student under a radiology resident's supervision. Test cases were segmented fully manually to avoid automation bias.

2.2 Pipeline Overview

Figure 1 illustrates the two-stage pipeline, consisting of (i) segmentation of BWT and (ii) classification of the underlying etiology.

For **Stage 1 (Segmentation)**, we employed nnU-Net [13], which automatically configures a 3D encoder–decoder based on dataset characteristics. CT volumes were cropped to the colon region using TotalSegmentator [22], resampled to 1 mm isotropic resolution, intensity-clipped to $[-100, 500]$ HU to maximize soft-tissue contrast, and min–max normalized. The model was trained via 5-fold cross-validation using a compound Dice and cross-entropy loss, with inference via sliding-window ensemble averaging.

For **Stage 2 (Classification)**, a 3D ResNet-18 [11] processes a dual-channel input: the CT region-of-interest (ROI) surrounding the predicted BWT concatenated with the corresponding binary segmentation mask. At inference, overlapping patches are evaluated via a sliding window and averaged into a single case-level score, enabling seamless classification over the entire ROI regardless of network input dimensions.

For further implementation details, code is available at: https://anonymous.4open.science/r/Differential-Diagnosis-of-Bowel-Wall-Thickening-on-CT-FBFA.

3 Results

3.1 Segmentation Performance

The nnU-Net achieved a global median Dice of 0.787 on the internal test set, with higher performance for colon cancer (median ≈0.83) than for diverticulitis (median ≈0.68). Qualitative examples in Fig. 2(d) demonstrate accurate spatial overlap between predictions and the reference standard, particularly in Cases 1–4. Cases 5–12 illustrate common failure modes, including incomplete capture of lesion morphology and spatial extent, failure to detect subtle lesions, and small false-positive clusters in adjacent regions.

3.2 Classification and Diagnostic Accuracy

The second-stage ResNet classifier demonstrated strong discriminative performance, achieving an area under the receiver operating characteristic curve (AUC) of 0.964 (Fig. 2(a)). On the internal test set, the model correctly classified 44 of 45 colon cancer cases and 26 of 32 diverticulitis cases (Fig. 2(b)). We conducted ablation studies on our internal dataset and two public external datasets. The first, the MSD-Colon dataset [1], contains 126 portal-venous phase CT scans of primary colon cancer. The second, the Stage-II-CC dataset [20], comprises 50 contrast-enhanced CT scans of histologically confirmed stage II colorectal cancer.

Pipeline ablation. Table 1 compares the proposed two-stage approach with two single-stage alternatives. A ResNet-18 applied directly to colon-cropped CT volumes achieved 0.75 accuracy and 0.41 specificity, indicating poor discrimination of diverticulitis despite a sensitivity of 1.00. A multi-label nnU-Net, trained to directly segment colon cancer and diverticulitis, achieved higher internal accuracy and specificity (both 0.94), but generalized poorly to external data and predicted both pathologies simultaneously in 43% of the cases. Case-level diagnosis was determined by majority vote over predicted pixels. Because segmentation optimizes voxel-level labeling rather than patient-level diagnosis, overlapping imaging patterns often produced ambiguous outputs. By separating lesion detection and diagnostic reasoning, our pipeline reduces this ambiguity and generalizes more robustly.

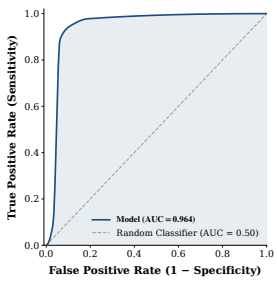
Input channel ablation. Adding the predicted segmentation mask as a second channel consistently improved classification over a CT-only baseline (Table 2). Internally, accuracy increased from 0.88 to 0.91 and AUC from 0.91 to 0.96. The AUC difference was statistically significant by the DeLong test ($p < 0.05$). The benefit was more pronounced on external data, with accuracy rising from 0.86 to 0.98 on MSD-Colon and from 0.90 to 0.96 on Stage-II-CC.

3.3 External validation and baseline comparison.

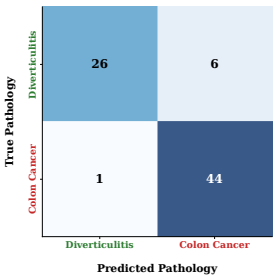
We evaluated the pipeline on an external mixed-etiology cohort from [ANONYMIZED] Hospital (120 patients, with 210 colon cancer and 214 diverticulitis CT scans across multiple scanners and contrast phases), achieving an accuracy of 0.63, sensitivity of 0.74, and specificity of 0.51. Because no prior work addresses this specific diagnostic task, we benchmarked four established 3D baselines (MedicalNet[3], SwinUNETR[10], DenseNet3D[12], and 3D-DINO[24]) trained on single-channel colon-cropped CT volumes. While these baselines underperformed our framework internally, they performed near chance (AUC $\leq$ 0.50) on the external [ANONYMIZED] cohort. This cross-architecture degradation underscores the severe covariate and domain shifts inherent to multicenter mixed-etiology data, suggesting that standard single-channel classification is inadequate for this task and supporting the use of a dual-channel localization strategy.

Table 1: Pipeline ablation. **Bold** = best per metric. ‡Multi-label model predicts both pathologies simultaneously in 43% of cases; majority-pixel label used as case-level decision.

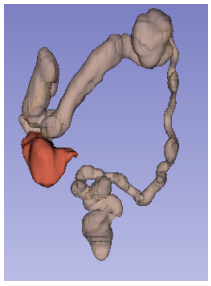
Pipeline	**Acc**	**Sens**	**Spec**	$\mathbf{Acc_{MSD}}$	$\mathbf{Acc_{S2}}$
2-stage: seg → cls *(ours)*	0.91	0.98	0.81	0.98	0.96
1-stage classification	0.75	**1.00**	0.41	**0.99**	**1.00**
1-stage segmentation‡	**0.94**	0.93	**0.94**	0.87	0.94



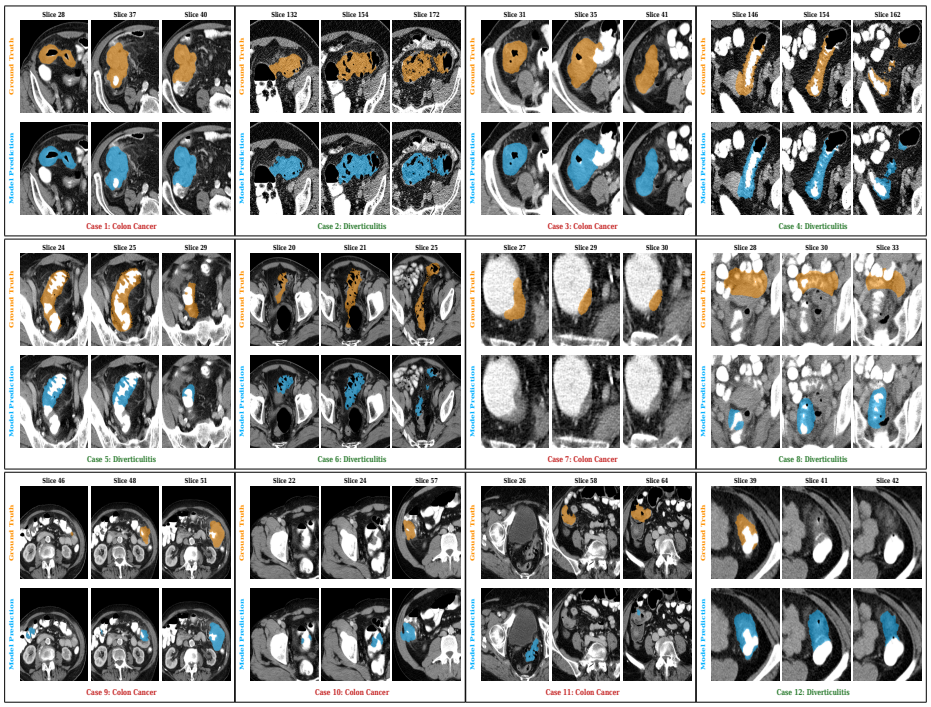
(a) ROC curve



(b) Confusion matrix



(c) 3D visualization



(d) Qualitative segmentation results

Fig. 2: **Overall segmentation and classification performance.** (a) ROC curve with an AUC of 0.964. (b) Confusion matrix for case-level classification. (c) Example 3D surface reconstruction of the colon segmented during preprocessing (brown) and the predicted BWT (red). (d) Qualitative segmentation results on representative CT volumes. Each box shows three axial slices (columns) with expert annotations (first row) and model predictions (second row).

Table 2: Input-channel ablation. Internal metrics report accuracy, sensitivity, specificity, and AUC (95% CI). External columns report accuracy on the cancer-only datasets. **Bold** indicates best performance.

Configuration	**Acc**	**Sens**	**Spec**	**AUC (95% CI)**	$\mathbf{Acc_{MSD}}$	$\mathbf{Acc_{S2}}$
Single (CT only)	0.88	0.96	0.78	0.910 (0.827, 0.971)	0.86	0.90
Dual (CT + mask)	**0.91**	**0.98**	**0.81**	**0.964** (0.895, 0.999)	**0.98**	**0.96**

Table 3: Classification performance of baseline models and the proposed method on the internal test set and external [ANONYMIZED] cohort. **Bold** indicates the best result per column.

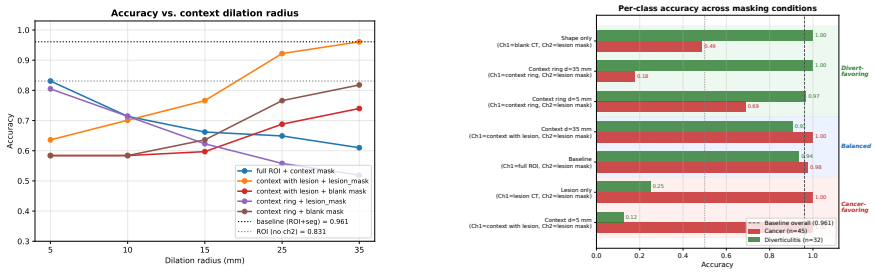
Model	**Internal**				**External**			
	Acc	Sens	Spec	AUC	Acc	Sens	Spec	AUC
DenseNet3D[12]	0.84	0.91	0.75	0.87	0.53	0.38	**0.67**	0.48
MedicalNet[3]	0.61	**1.00**	0.06	0.60	0.47	**0.87**	0.07	0.45
SwinUNETR[10]	0.71	0.87	0.50	0.67	0.46	0.62	0.29	0.47
3D-DINO[24]	0.74	0.97	0.41	0.82	0.51	0.48	0.54	0.50
Ours	**0.91**	**0.98**	**0.81**	**0.96**	**0.63**	**0.74**	0.51	**0.67**

3.4 Classifier Decision Criteria

To probe which image regions drive classification, we performed input masking experiments using the model trained with reference lesion segmentations. Channel 1 (Ch1) contained CT intensities and Channel 2 (Ch2) the BWT mask. We varied the visible CT region in Ch1 (lesion, lesion + context, context ring, full ROI, or none) and the mask input in Ch2 (lesion, dilated, or removed). Context extent was controlled by isotropic dilation of the lesion mask with radii $d \in [5, 35]$ mm.

The per-class analysis (Figure 3b) showed asymmetric reliance on distinct image features for each diagnosis. Cancer classification relied strongly on intra-lesional appearance. Accuracy was 1.00 whenever lesion pixels were visible in Ch1, regardless of context extent, but decreased markedly when the lesion was removed, dropping to 0.18 under the Ring $d = 35$ mm condition and to 0.49 under the shape-only condition. Conversely, diverticulitis classification relied almost exclusively on shape and surrounding tissue. Accuracy was low with lesion-only input (0.25) but reached near-perfect levels under shape-only input (1.00) and the ring conditions (0.97–1.00), where internal lesion content was omitted.

The dilation sweep (Figure 3a) confirmed these dynamics across expanding radii. When the true lesion mask was provided alongside the context (orange line), overall accuracy increased monotonically from 0.64 to 0.96 as the visible context grew, confirming that pericolic tissue provides critical complementary information. Removing the lesion pixels while retaining the segmentation mask (purple line) degraded accuracy substantially, declining to approximately 0.52 at larger radii, underscoring that lesion tissue is required for high-confidence classification. When the segmentation mask was entirely omitted (red and brown



(a) (b)

Fig. 3: **Input masking analysis(classifier trained with reference lesion segmentations; Ch1 = ROI CT, Ch2 = lesion segmentation).** (a) Overall accuracy as a function of context dilation radius (d) for different input masking schemes. (b) Per-class accuracy for cancer (red) and diverticulitis (green)

lines), accuracy increased with radius but plateaued well below the baseline in both cases, demonstrating that surrounding anatomy contains genuine diagnostic signal but remains insufficient without a localization prior. Finally, replacing the true lesion mask with a coarse, expanding context mask (blue line) produced a monotonic decline from 0.831 to approximately 0.61. This suggests that Ch2 does not act as a generic spatial pointer. Rather, precise lesion delineation is required to unlock the model's discriminative capacity.

4 Discussion

The proposed fully automated two-stage pipeline, consisting of bowel wall thickening (BWT) segmentation followed by dual-channel classification, eliminates the need for manual ROI selection required by the most comparable prior work [26], while achieving a malignancy sensitivity of 0.98 and an AUC of 0.964 on the internal test set. Segmentation performance remained competitive despite the greater complexity of mixed-etiology BWT compared with conventional tumor-only segmentation[17,25,23,9,8]. Overall Dice reached 0.787, with lower performance for diverticulitis (median 0.68) than colon cancer (0.83), likely reflecting the diffuse morphology, ill-defined boundaries, and annotation uncertainty characteristic of inflammatory disease. A board-certified radiologist's review of lower-scoring segmentations showed that voxel-wise metrics often underestimate clinical utility and revealed four recurring discrepancy patterns. First, in Cases 5–6, the model correctly localized disease but was penalized for partial boundary misdelineation. Second, in Case 12, the model identified regions not captured by the incomplete reference standard, resulting in predictions extending beyond the annotations. Third, apparent false positives (Cases 8–11) frequently represented clinically plausible but unannotated findings attributable to diverticulitis, including atypical inflammatory changes, physiological peristalsis artifacts, and separate areas of chronic diverticulitis. Fourth, true model failures were rare and

limited to empty segmentations in very mild disease (Case 7). Overall, disagreements primarily reflected annotation incompleteness and boundary ambiguity rather than failure to detect clinically relevant BWT.

Pipeline ablation and external validation both support the value of explicit localization. The one-stage classification baseline showed substantially lower specificity (0.41), while the multi-label segmentation approach frequently produced simultaneous predictions for both diagnoses, limiting clinical usability. Externally, although performance decreased on the heterogeneous [ANONYMIZED] cohort (accuracy 0.63, sensitivity 0.74, specificity 0.51), malignancy sensitivity remained relatively preserved, whereas all other baselines performed near chance. Finally, the input masking experiments (Fig. 3) provide mechanistic insight into model behavior. Cancer classification depended predominantly on intra-lesional appearance, whereas diverticulitis remained highly classifiable from lesion morphology and perilesional context alone. Performance improved with increasing contextual information only when accurate lesion localization was preserved, indicating that the segmentation channel functions as a meaningful structural prior rather than a generic attention mechanism. The model therefore appears to learn radiologically plausible decision pathways: lesion-centered appearance cues for cancer and contextual inflammatory features for diverticulitis.

This study has some limitations. Reference annotations for diffuse BWT are inherently uncertain, and residual interobserver variability likely limits the achievable segmentation performance. Although the cancer-only external datasets demonstrated robust generalization for malignancy detection, the results on the [ANONYMIZED] cohort indicate that further improvements in robustness and cross-center generalizability are needed.

5 Conclusion

We present the first fully automated framework for CT-based differential diagnosis of bowel wall thickening, combining lesion localization with dual-channel classification to distinguish colon cancer from diverticulitis. The proposed two-stage pipeline achieved high malignancy sensitivity (0.98) and strong internal discrimination while generalizing to external cancer datasets and maintaining clinically relevant cancer detection under substantial multicenter distribution shift. Input masking further demonstrated radiologically plausible decision behavior, showing complementary reliance on lesion appearance and perilesional context. Future work will focus on improving generalizability through multi-institutional training and extending the framework to additional causes of BWT.

Prospects of Application. This fully automated pipeline is well suited for integration into emergency radiology and high-throughput triage workflows, where speed and reproducibility are critical. Operating as an objective second reader, it could help reduce two costly diagnostic failures, delayed cancer detection and unnecessary invasive workup, and direct patients toward appropriate surgical or conservative management pathways at initial CT presentation without requiring workflow redesign.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M.: The Medical Segmentation Decathlon. Nature Communications **13**(1), 4128 (2022)
2. Azhar, N., Buchwald, P., Ansari, H., Schyman, T., Yaqub, S., Øresland, T., Schultz, J.: Risk of Colorectal Cancer Following CT-Verified Acute Diverticulitis: A Nationwide Population-Based Cohort Study. Colorectal Disease **22**(10), 1406–1414 (2020)
3. Chen, Sihong and Ma, Kai and Zheng, Yefeng: Med3D: Transfer Learning for 3D Medical Image Analysis. arXiv Preprint arXiv:1904.00625 (2019)
4. Chintapalli, K.N., Chopra, S., Ghiatas, A.A., Esola, C.C., Fields, S.F., Dodd III, G.D.: Diveriticulitis Versus Colon Cancer: Differentiation With Helical CT Findings. Radiology **210**(2), 429–435 (1999)
5. Fernandes, T., Oliveira, M.I., Castro, R., Araújo, B., Viamonte, B., Cunha, R.: Bowel Wall Thickening at CT: Simplifying the Diagnosis. Insights into Imaging **5**(2), 195–208 (2014)
6. Fogelstrom, A., Hallen, F., Pekkari, K.: Computed Tomography Diagnosed First Time Diverticulitis and Colorectal Cancer. International Journal of Colorectal Disease **35**(10), 1895–1901 (2020)
7. George, N., Khan, S., Gunes, A., Machado, P., Bade-Boon, J., Badov, D.: The Significance of Bowel Wall Thickening on Abdominal Computed Tomography at an Australian Hospital: 163. Official Journal of The American College of Gastroenterology| ACG **111**, S77–S78 (2016)
8. Ghosal, S., Das, D., Rai, J.K., Pandaw, A.S., Verma, S.: iScan: Detection of Colorectal Cancer from CT Scan Images Using Deep Learning. ACM Trans. Comput. Healthcare **5**(3) (Sep 2024). https://doi.org/10.1145/3676282, https://doi.org/10.1145/3676282
9. Hassan, G.M., Datta, A.: 3D-SCUMamba: An Abdominal Tumor Segmentation model. IEEE Access (2025)
10. Hatamizadeh, Ali and Nath, Vishwesh and Tang, Yucheng and Yang, Dong and Roth, Holger R. and Xu, Daguang: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2021)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition (2015), https://arxiv.org/abs/1512.03385
12. Huang, Gao and Liu, Zhuang and van der Maaten, Laurens and Weinberger, Kilian Q.: Densely Connected Convolutional Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708 (2017)
13. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., Maier-Hein, K.H.: nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation (2018), https://arxiv.org/abs/1809.10486
14. Koo, C.H., Chang, J.H., Syn, N.L., Wee, I.J., Mathew, R.: Systematic Review and Meta-Analysis on Colorectal Cancer Findings on Colonic Evaluation After CT-Confirmed Acute Diverticulitis. Diseases of the Colon & Rectum **63**(5), 701–709 (2020)
15. Lau, K., Spilsbury, K., Farooque, Y., Kariyawasam, S., Owen, R., Wallace, M., Makin, G.: Is Colonoscopy Still Mandatory After A CT Diagnosis of Left-Sided Diverticulitis: Can Colorectal Cancer Be Confidently Excluded? Diseases of The Colon & Rectum **54**(10), 1265–1270 (2011)

16. Meyer, J., Orci, L.A., Combescure, C., Balaphas, A., Morel, P., Buchs, N.C., Ris, F.: Risk of Colorectal Cancer in Patients With Acute Diverticulitis: A Systematic Review and Meta-Analysis of Observational Studies. Clinical Gastroenterology and Hepatology **17**(8), 1448–1456 (2019)
17. Pei, Y., Mu, L., Fu, Y., He, K., Li, H., Guo, S., Liu, X., Li, M., Zhang, H., Li, X.: Colorectal Tumor Segmentation of CT Scans Based on a Convolutional Neural Network With an Attention Mechanism. IEEE Access **8**, 64131–64138 (2020)
18. Sahoo, P.K., Gupta, P., Lai, Y.C., Chiang, S.F., You, J.F., Onthoni, D.D., Chern, Y.J.: Localization of Colorectal Cancer Lesions in Contrast-Computed Tomography Images via a Deep Learning Approach. Bioengineering **10**(8) (2023). https://doi.org/10.3390/bioengineering10080972, https://www.mdpi.com/2306-5354/10/8/972
19. Sai, V.F., Velayos, F., Neuhaus, J., Westphalen, A.C.: Colonoscopy After CT Diagnosis of Diverticulitis to Exclude Colon Cancer: A Systematic Literature Review. Radiology **263**(2), 383–390 (2012)
20. Tong, T., Li, M.: Abdominal or Pelvic Enhanced CT Images Within 10 Days Before Surgery of 230 Patients With Stage II Colorectal Cancer (StageII-Colorectal-CT). The Cancer Imaging Archive (2022). https://doi.org/10.7937/p5k5-tg43, https://doi.org/10.7937/p5k5-tg43
21. Wan, S., Guo, W., Zou, B., Wang, W., Qiu, C., Liu, K., Jin, P., Yang, J.: Tuning Vision Foundation Models for Rectal Cancer Segmentation from CT Scans: Development and Validation of U-SAM (2024)
22. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiology: Artificial Intelligence **5**(5) (Sep 2023). https://doi.org/10.1148/ryai.230024, http://dx.doi.org/10.1148/ryai.230024
23. Xie, J., Zhang, Z., Luo, G., Zhu, Y.: A Segmentation Foundation Model for Diverse-type Tumors. arXiv preprint arXiv:2403.06396 (2024)
24. Xu, Tony and Hosseini, Sepehr and Anderson, Chris and Rinaldi, Anthony and Krishnan, Rahul G. and Martel, Anne L. and Goubran, Maged: A Generalizable 3D Framework and Model for Self-Supervised Learning in Medical Imaging. npj Digital Medicine **8**(1), 639 (2025)
25. Zhang, R., Bai, Z., Yu, R., Pang, W., Wang, L., Zhu, L., Zhang, X., Zhang, H., Hu, W.: AG-CRC: Anatomy-Guided Colorectal Cancer Segmentation in CT with Imperfect Anatomical Knowledge. arXiv preprint arXiv:2310.04677 (2023)
26. Ziegelmayer, S., Reischl, S., Havrda, H., Gawlitza, J., Graf, M., Lenhart, N., Nehls, N., Lemke, T., Wilhelm, D., Lohöfer, F.: Development and Validation of a Deep Learning Algorithm to Differentiate Colon Carcinoma From Acute Diverticulitis in Computed Tomography Images . JAMA Network Open **6**(1), e2253370–e2253370 (2023)