

# Artificial Intelligence and Radiomics for Detection, Segmentation, and Biomarker Prediction in Gastrointestinal Cancers

Niloofer Hooshmand<sup>1#</sup>, Mobin Ahmadi<sup>1#</sup>, Ramin Raoufinia<sup>2,3#</sup>, Aida Yavari Kondori<sup>1</sup>, Ibrahim Saeed Gataa<sup>4</sup>, Ali Gholami<sup>2,3</sup>, Amir Avan<sup>1\*</sup>, Hamid Naderi<sup>1\*</sup>

1- Metabolic Syndrome Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

2- Healthy Ageing Research Centre, Neyshabur University of Medical Sciences, Neyshabur, Iran.

3- Noncommunicable Diseases Research Center, Neyshabur University of Medical Sciences, Neyshabur, Iran.

4- College of Medicine, University of Warith Al-Anbiyaa, Karbala 56001, Iraq.

+ These authors contributed equally and share first authorship.

**\*Corresponding authors:** Amir Avan, and Hamid Naderi; Metabolic Syndrome Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. E-mail: [avana@mums.ac.ir](mailto:avana@mums.ac.ir), [amir\\_avan@yahoo.com](mailto:amir_avan@yahoo.com)

**Received** 2026 July 23; **Accepted** 2026 September 25.

## Abstract

**Background:** Early diagnosis of gastrointestinal (GI) cancers remains challenging because of the heterogeneity of imaging modalities and clinical manifestations and the absence of specific symptoms. Artificial intelligence (AI), deep learning, and radiomics can enhance image interpretation, automate tumor segmentation, and identify clinically relevant imaging biomarkers.

**Objectives:** This narrative review aimed to summarize recent advances in AI- and radiomics-based imaging of colorectal, gastric, liver, and pancreatic cancers, focusing on tumor detection, segmentation, staging, histological classification, molecular biomarker prediction, and prognostic evaluation.

**Materials and Methods:** A structured, topic-focused search of PubMed/MEDLINE and Google Scholar was conducted without a lower publication-date limit and was last updated on 30 July 2026. English-language human studies evaluating AI- or radiomics-based computed tomography (CT), magnetic resonance imaging, or positron emission tomography applications were considered. Studies with external validation, multicenter designs, or larger cohorts, as well as systematic reviews and meta-analyses, were prioritized, and methodological robustness was appraised qualitatively.

**Results:** In colorectal cancer, AI-based methods have been developed for preoperative staging, prediction of Kirsten rat sarcoma viral oncogene homolog (KRAS) mutation and microsatellite instability, lesion segmentation, and survival estimation. In gastric cancer, deep learning and radiomics have been applied to cancer detection on noncontrast CT, cross-center three-dimensional tumor segmentation, and preoperative histological type prediction. Liver cancer studies have mostly focused on automated lesion segmentation and radiomics-based prediction of clinically relevant biomarkers such as vessels encapsulating tumor clusters. In pancreatic cancer, multimodal and attention-based models have been used for three-dimensional and small-tumor segmentation and for prediagnostic detection of pancreatic ductal adenocarcinoma. Although several models achieved high area-under-the-curve or Dice values, direct comparisons are limited by heterogeneity in populations, imaging protocols, model architectures, performance metrics, and validation strategies. Much of the evidence remains retrospective and is often based on single-center or selected datasets with limited external validation.

**Conclusions:** AI and radiomics show potential as decision-support tools in GI cancer imaging. Before clinical application, prospective multicenter evaluation, standardized imaging and reporting practices, and improved accessibility and integration into routine radiology workflows are required.

**Keywords:** Gastrointestinal Cancer; Artificial Intelligence; Radiomics; Deep Learning; Early Detection; Oncologic Imaging.

## 1. Introduction

Gastrointestinal (GI) cancers include malignancies of the esophagus, stomach, small intestine, colorectum, anus, liver, intrahepatic and extrahepatic bile ducts, gallbladder, and pancreas (1). GI cancers

account for approximately 26% of global cancer incidence and 35% of all cancer-related deaths (2). Each year, over 1.9 million new patients are diagnosed with colorectal cancer (CRC) worldwide, making it the third most common cancer and a major global cause of cancer-related mortality (3). In CRC,

metastatic disease is associated with substantially poorer survival and increased mortality (4-6). Gastric cancer is the fourth leading cause of cancer death and the fifth most commonly diagnosed cancer worldwide (7). Most gastric-cancer cases and deaths occur in Asia, with a particularly high burden in Eastern Asia (8, 9). As another important global health concern, liver cancer is the sixth most frequently diagnosed malignancy and the third most common cause of cancer mortality, after lung and colorectal malignancies (3). Hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC) are the most frequently occurring types of primary liver cancer (10). Chronic hepatitis B virus (HBV) infection is an important etiologic factor for HCC and may lead to cirrhosis or hepatocellular carcinoma (10-12). The Liver Imaging Reporting and Data System (LI-RADS) standardizes liver-imaging terminology and reporting for HCC surveillance, diagnosis, and treatment-response assessment (13). In a selected cohort of 163 patients with mass-forming intrahepatic cholangiocarcinoma who underwent microscopically margin-negative (R0) resection and had accurate pathological nodal staging, median survival was 53 months for American Joint Committee on Cancer (AJCC) seventh-edition stage II and 16 months for stage III (14). Although gastrointestinal cancers encompass a broad group of malignancies, this review focuses on four major cancer types within the predefined imaging-based Artificial Intelligence (AI) and radiomics scope: colorectal, gastric, liver, and pancreatic cancers.

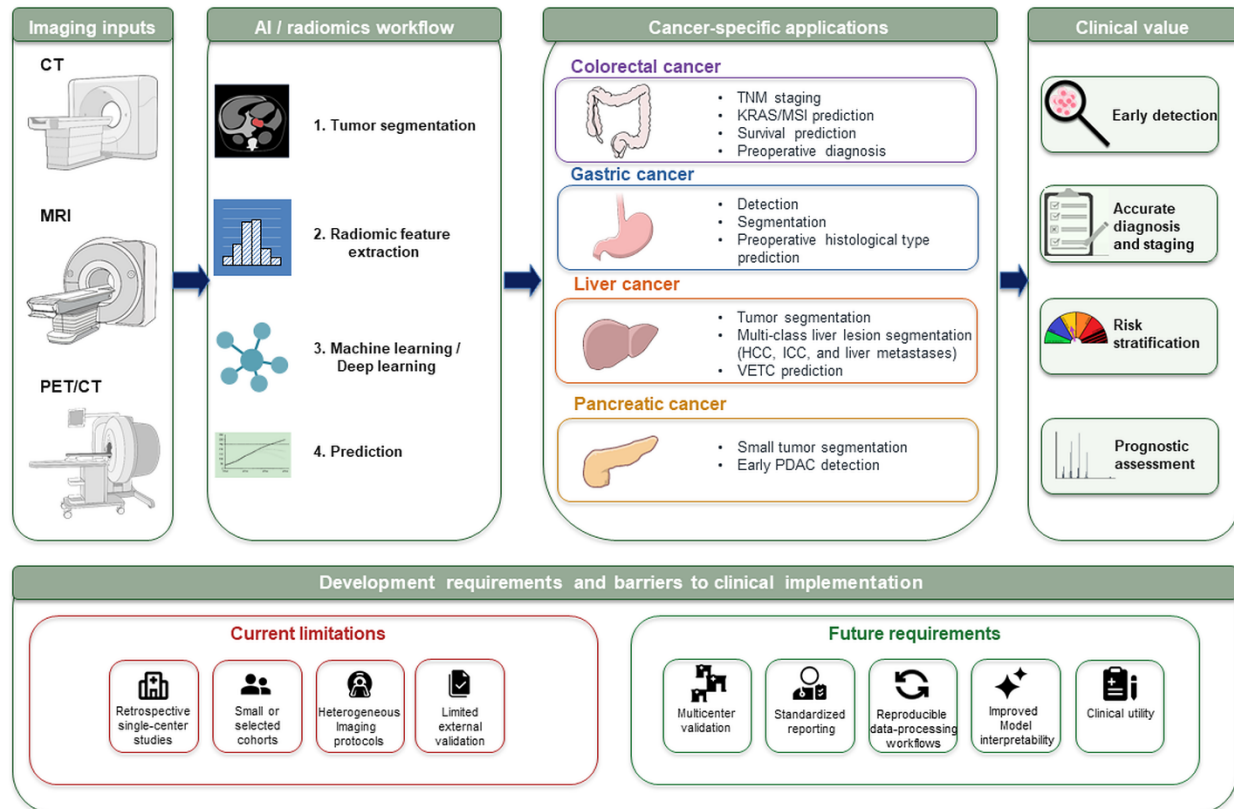
Imaging plays a crucial role in evaluating GI cancers and determining appropriate treatment approaches. Pelvic magnetic resonance imaging (MRI) is recommended for the primary staging of rectal cancer and is used to guide clinical management (15).

Although these imaging methods have

advanced considerably, manual evaluation of imaging findings is time-consuming and subject to inter-observer variability that can affect diagnostic accuracy. The need for consistent, rapid, and automated techniques in medical imaging has thus resulted in a wide range of research on AI and deep learning (DL) (16, 17). DL approaches have been widely applied to medical image analysis, including image classification, object detection, and segmentation. For instance, Convolutional Neural Networks (CNNs) have been widely applied for medical image segmentation, including patch-based methods that classify each pixel based on information from its surrounding image areas (16).

Radiomics is a rapidly developing field that extracts large amounts of quantitative data from medical images, making them suitable for analysis (18). Radiomic features may correlate with genomic characteristics and capture tumor heterogeneity across the entire tumor volume, thereby supporting prognostic modeling and patient stratification (19). Radiomics models derived from CT and MRI have been investigated as non-invasive tools for preoperative prediction of microsatellite instability (MSI) status in CRC (20). MSI is now recognized as a key biomarker for immunotherapy and chemotherapy response and serves as a prognostic marker in CRC. Standard MSI assessment relies on polymerase chain reaction (PCR) or immunohistochemistry performed on tissue samples, which may be affected by sampling limitations and intratumor heterogeneity (20).

Therefore, this review highlights key AI-driven advances in radiological detection and diagnostic evaluation of major GI malignancies, discusses existing challenges in clinical application, and highlights future perspectives for integrating AI into routine oncologic imaging practice. (Figure 1).



**Figure 1.** General workflow of artificial intelligence (AI) and radiomics and their major clinical applications in gastrointestinal cancer imaging. Medical images from CT, MRI, and PET/CT are analyzed using AI- and radiomics-based workflows, including tumor segmentation, radiomic feature extraction, machine-learning and deep-learning analysis, and clinical prediction. Representative applications include tumor detection, segmentation, staging, molecular biomarker prediction, histological classification, prognostic assessment, and early cancer detection in colorectal, gastric, liver, and pancreatic cancers. The key challenges for clinical implementation and the main requirements for future clinical translation are demonstrated.

**Abbreviations:** AI, artificial intelligence; CT, computed tomography; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; KRAS, Kirsten rat sarcoma viral oncogene homolog; MRI, magnetic resonance imaging; MSI, microsatellite instability; PDAC, pancreatic ductal adenocarcinoma; PET/CT, positron emission tomography/computed tomography; TNM, tumor–node–metastasis; VETC, vessels encapsulating tumor clusters.

## 2. Literature Search and Study Selection

A structured, topic-focused search of PubMed/MEDLINE and Google Scholar was conducted without a lower publication-date limit and was last updated on 30 July 2026. The search covered colorectal, gastric, liver, and pancreatic cancers; computational approaches, including artificial intelligence, machine learning, deep learning, and radiomics; imaging modalities, including CT, MRI, and positron emission tomography (PET); and clinical applications, including detection, segmentation, staging, biomarker prediction, and prognosis. Terms were combined using the Boolean operators AND

and OR. English-language human studies evaluating AI- or radiomics-based imaging applications were considered. Nonhuman studies, non-English publications, and articles unrelated to the predefined cancer types or imaging applications were excluded. When available, studies with external validation, multicenter designs, larger cohorts, systematic reviews, and meta-analyses were prioritized because of their potential contribution to methodological robustness and clinical relevance. Relevant reference lists were also screened to identify additional studies. Because this review was narrative in nature, the search and selection process was not intended to provide quantitative

estimates of study flow; therefore, numerical counts of retrieved records, full-text evaluations, and included studies were not used as selection metrics. The identified studies were further evaluated based on their relevance to the objectives of this review, methodological contribution, clinical significance, and ability to represent key advances and challenges in AI- and radiomics-based gastrointestinal cancer imaging. Reference selection was performed through an iterative qualitative assessment rather than a quantitative screening process. Because this article was designed as a narrative rather than a systematic review, no formal systematic-review protocol, numerical risk-of-bias scoring, or quantitative meta-analysis was performed. Given the methodological heterogeneity of the included evidence, which encompassed radiomics, machine learning, end-to-end deep learning, classification, segmentation, biomarker prediction, and multimodal imaging approaches, a single post hoc numerical quality score was not applied across studies. Instead, the methodological robustness of the principal benchmark studies was appraised qualitatively and comparatively, with emphasis on cohort size and setting, internal versus external testing, imaging and preprocessing heterogeneity, generalizability, and study-specific methodological limitations. For Table 1, external validation was defined as evaluation in an independent cohort not used for model development, preferably from a different institution. Cross-centre domain-adaptation testing, public-benchmark evaluation without a clearly independent test set, pooled meta-analytic validation, and real-world operational evaluation were reported separately when independence was not established.

## 2.1. Clinical Applications of AI in Colorectal Cancer Imaging

Colorectal cancer (CRC) ranks among the

most common malignancies globally (21). Compared to no endoscopic screening, screening colonoscopy was associated with a 67% lower risk of death from any colorectal cancer (22). However, the current screening methods are not widely adopted due to multiple systemic and procedural barriers (23). CT colonography is a minimally invasive imaging method that has been evaluated against colonoscopy in prospective studies (24-28). In a 15-center study with complete data from 2,531 asymptomatic adults, the per-patient sensitivity of CT colonography was 0.78 for adenomas  $\geq 6$  mm and 0.90 for adenomas or cancers  $\geq 10$  mm (26). Computer-aided detection (CAD) systems have been developed to assist lesion detection in CT colonography (29, 30). Moreover, accurate determination of Kirsten rat sarcoma viral oncogene homolog (KRAS) mutation status is important for tumor prognosis and personalized treatment in patients with CRC (31). More broadly, colorectal tumorigenesis involves dysregulation of interconnected oncogenic signaling networks, including the PI3K/AKT/mTOR and Wnt/ $\beta$ -catenin pathways (32).

Yang et al. showed that the CRCFound\_CT model was evaluated across eight clinically relevant diagnostic and prognostic tasks, including tumor-node-metastasis (TNM) staging, which assesses local tumor invasion (T), lymph-node involvement (N), and distant metastasis (M). The model also showed good discrimination in predicting microsatellite instability. In the external test dataset, CRCFound\_CT achieved average areas under the curve (AUCs) of 0.7623, 0.827, and 0.7051 for T-, N-, and M-stage prediction, respectively. However, performance was lower for more complex cases, suggesting that additional refinement of the model is required (33). In contrast, the Patch Embedding Dual Residual Attention Mechanism-EfficientB0 (PEDRA-EFB0) model was designed for CT-based feature extraction

and survival prediction in colorectal cancer. Zhao et al. integrated patch embeddings and a dual residual attention mechanism to enhance feature extraction from CRC CT images, and the implementation code was made freely available (34).

CRCFound was developed using data from 6,332 patients, whereas PEDRA-EFBO included 801 patients; however, differences in study purpose, outcome definition, imaging protocol, and validation design preclude direct performance comparison between the two models (33, 34).

AI has also been developed for KRAS mutation identification. Song et al. developed a multi-task, attention-based deep learning framework that uses T2-weighted MRI to identify it. The framework combines lesion images with segmentation results to extract additional features and uses an interlayer loss function to alleviate, to some extent, the overfitting associated with the small dataset. The mean accuracy for determining KRAS mutation status was  $89.95 \pm 1.23\%$ , the mean sensitivity was  $89.29 \pm 1.79\%$ , the mean specificity was  $90.53 \pm 2.45\%$ , and the mean AUC was  $95.73 \pm 0.52\%$ . For lesion segmentation, the mean Dice value was  $88.11 \pm 0.86\%$ . The model was evaluated primarily using 15-fold cross-validation on a relatively small dataset comprising 382 images from 36 clinical cases (31). Therefore, its reported performance requires confirmation in larger independent cohorts.

In this regard, a systematic review identified 29 radiomic studies on KRAS mutation prediction in CRC, 19 of which were included in the meta-analysis. The training cohorts' pooled sensitivity, specificity, and AUC were 0.80 (95% confidence interval (CI), 0.75–0.84), 0.80 (95% CI, 0.74–0.85), and 0.87 (95% CI, 0.84–0.90), respectively. For the validation cohorts (13 studies), the pooled sensitivity, specificity, and AUC were 0.78 (95% CI, 0.71–0.84), 0.84 (95% CI, 0.74–

0.90), and 0.86 (95% CI, 0.83–0.89), respectively. Consequently, radiomic models have the potential to differentiate KRAS status in colorectal cancer (CRC). However, they are still in the early stages and need higher methodological quality to demonstrate their feasibility and clinical applicability. Accordingly, radiomics models for KRAS prediction remain investigational and require further prospective and external validation before clinical implementation (35). Radiogenomic associations should be interpreted within a biological rather than purely statistical framework. Molecular alterations may first reshape tissue architecture and the tumor microenvironment, including vascular/perfusion and extracellular-matrix remodeling, and these changes may subsequently appear on imaging as differences in tumor morphology, intensity, and texture. However, such radiogenomic associations should not be interpreted as proof that an individual radiomic feature directly represents a specific molecular pathway (36). This distinction is particularly important in small cohorts because insufficient sample size may increase the risk of overfitting and overestimation of model performance (37). Accordingly, the findings of Song et al., based on only 36 clinical cases and primarily evaluated using 15-fold cross-validation, require confirmation in larger independent cohorts (31).

## 2.2. Clinical Applications of AI in Gastric Cancer Imaging

Gastric cancer (GC) is one of the most common human malignancies worldwide (7). Several imaging techniques are used for GC diagnosis. Noncontrast CT examinations are widely performed for other clinical indications and may be leveraged for opportunistic gastric-cancer detection, including in settings with limited screening resources (38, 39). Endoscopic ultrasonography (EUS) is commonly used for

locoregional staging of gastric cancer but is operator-dependent and invasive (40). Contrast-enhanced CT plays an important role in gastric cancer staging, including assessment of the primary tumor, regional lymph nodes, and distant metastases (40).

Nevertheless, for AI-based opportunistic screening on routine CT, results must still be shown to generalize across scan acquisition parameters, reconstruction settings, scanner vendors, and imaging centers before broader clinical adoption (39). Hu et al. proposed the Gastric Cancer Risk Assessment Procedure with AI (GRAPE), a deep learning-based approach that utilizes noncontrast CT scans to detect patients with gastric cancer. GRAPE outperformed earlier gastric cancer risk models based on clinical or serological data, including models with reported AUCs of 0.757 and 0.79 (38, 41, 42). The GRAPE sensitivity in subgroup analysis was about 50% for early gastric cancer and >90% for gastric cancer in stages T3 and T4. Prospective evaluation and further improvement in sensitivity for early gastric cancer are required before population-level screening implementation (38).

In addition to cancer detection, AI has been developed for automated tumor segmentation. Yuan et al. proposed a cross-center three-dimensional tumor segmentation model, Hierarchical Class-Aware Domain-Adaptive Network (HCA-DAN), to extract multi-scale features from anisotropic CT images and minimize the performance degradation induced by inter-center domain shifts. The Anisotropic neural network and Transformer (AsTr) baseline achieved Dice similarity coefficients (DSCs) of 59.26%, 55.97%, 48.83%, and 67.28% in four in-center tasks and 56.42%, 55.94%, 46.54%, and 60.62% in four cross-center tasks. HCA-DAN improved the cross-center DSCs to 58.36%, 56.72%, 49.25%, and 62.20%, respectively. The method requires access to target-domain images and center-specific

adaptation, which may limit direct deployment in previously unseen clinical settings (43). Furthermore, Huang et al. developed a CT-based radiomics nomogram that integrates selected radiomics features with cancer antigen 125 (CA125) for the preoperative prediction of gastric cancer histological types. The AUC values of the CT-based radiomics nomogram in the training, internal validation, and external validation groups were 0.777, 0.726, and 0.720, respectively, indicating its acceptable discriminative performance in preoperative histological classification of gastric cancer. However, their sample size was small, and radiomics feature extraction was limited to CT images in the portal vein phase, limiting the potential of their method (44).

Consequently, AI-based approaches have shown potential for gastric-cancer detection and preoperative histological classification. Among the reviewed gastric cancer studies, GRAPE was evaluated in the largest population, including model development using data from two centers, independent external validation across 16 centers, and real-world evaluation in three hospitals. However, its approximately 50% sensitivity for early gastric cancer remains an important limitation for screening applications (38).

### 2.3. Clinical Applications of AI in Liver Cancer Imaging

Atypical CT and MRI appearances of hepatocellular tumors can create diagnostic challenges and lead to misinterpretation (45). Accurate liver and tumor segmentation can support lesion assessment, treatment planning, and longitudinal evaluation (46). Manual segmentation of liver lesions is time-consuming and subject to interobserver variability. Therefore, developing automated methods for liver tumor segmentation is important for supporting clinical image analysis. However, automatic liver tumor segmentation remains challenging because of large variations in tumor size and location

(47), and ambiguous boundaries between tumor and normal tissue (48).

Xie et al. demonstrated that the performance of deep learning models for less frequent classes can be negatively impacted by an imbalance of liver cancer types in the training dataset. Their approach was trained on CT images and associated tumor segmentations from 591 patients with hepatocellular carcinoma, intrahepatic cholangiocarcinoma, or liver metastases. The authors proposed an adaptive balanced Dice loss to improve feature learning among tumor classes and combined it with a greedy parameter averaging strategy to allow optimization. On the test set, the model achieved a Dice-per-case of 0.819 (95% CI, 0.798-0.841), volumetric overlap error of 0.324 (95% CI, 0.268-0.380), and average symmetric surface distance of 4.69 (95% CI, 3.10-6.27). However, the absence of benign lesions in the evaluated cohort may limit the generalizability of the model to more heterogeneous clinical datasets (49).

Singh et al. proposed FasNet, a hybrid deep-learning framework based on VGG-16 and ResNet-50 for liver tumor segmentation. VGG-16 was utilized to capture spatial fine details, and ResNet-50 was used to discover higher-level hierarchical features. Modules for channel and spatial attention were added to focus on the most informative features and spatial regions. On the LiTS17 dataset, FasNet obtained a Dice coefficient of 0.8766 and a Jaccard index of 0.8487. Larger and more diverse datasets, including cohorts based on MRI, are required for further validation before clinical implementation (50).

Furthermore, Wu et al. proposed MSA-UNet, a CT-based segmentation model that combines U-Net architecture with multiscale feature extraction and an attention mechanism. On the test set, MSA-UNet achieved a DSC of  $98.00 \pm 0.38\%$ , an intersection over union of  $96.08 \pm 0.74\%$ , a

precision of  $97.17 \pm 0.85\%$ , and a recall of  $98.85 \pm 0.70\%$  for liver organ segmentation, outperforming U-Net, U-Net++, DeepLabv3+, and CE-Net across all four-evaluation metrics. However, further evaluation using larger datasets and development of a more lightweight network with fewer parameters are required (51).

Among segmentation studies, Xie et al. (49) analyzed the most heterogeneous cohort, while the framework proposed by Wu et al. (51) demonstrated higher DSC than Singh et al. (50). However, those studies were evaluated with a different segmentation task. Generally, organ segmentation is an easier problem than lesion-level or multi-class tumor segmentation. Therefore, differences in the DSC values can be affected by the dataset differences, segmentation task, and the validation protocols.

Besides tumor segmentation, AI has recently been employed in radiomics-based prediction of clinically relevant biomarkers in HCC. A systematic review and meta-analysis evaluated the diagnostic accuracy and prognosis of radiomics-based AI models for predicting Vessels Encapsulating Tumor Clusters (VETC) status in HCC. Contrast-enhanced magnetic resonance imaging (CEMRI)-based AI models showed the highest sensitivity (0.84) and AUC (0.87) for predicting VETC in HCC, whereas the single CEUS-based study showed the highest specificity (0.81), according to a meta-analysis of 15 studies. However, the generalizability of the findings was limited because 14 of the 15 included studies were conducted in China. Furthermore, most studies had small sample sizes (52). The biological plausibility of imaging-based VETC prediction is further supported by recent multicenter radiopathomics evidence showing that VETC represents a distinct vascularization pattern in HCC, histologically characterized by sinusoid-like CD34-positive

vessels encapsulating tumor clusters, and that deep radiomics models derived from gadoteric acid-enhanced MRI predicted VETC in an external test set with AUCs of 0.77–0.79 (53). Consequently, AI models have shown progress in liver and tumor segmentation and in imaging-based prediction of clinically relevant biomarkers. However, larger and more diverse cohorts, standardized validation, and external multicenter evaluation are required before clinical implementation.

#### 2.4. Clinical Applications of AI in Pancreatic Cancer Imaging

Pancreatic cancer (PC) is one of the most life-threatening cancers (54). Early clinical manifestations of pancreatic cancer are often nonspecific, and no individual initial symptom reliably distinguishes pancreatic cancer from non-cancer conditions. However, subsequent symptoms such as jaundice, fatigue, changes in bowel habits, weight loss, and decreased appetite may signal the disease (55). Therefore, earlier diagnosis remains important for improving the opportunity for potentially curative treatment (56, 57). AI-assisted methods are increasingly being investigated for pancreatic-cancer imaging (58).

A hierarchical multimodal hybrid encoding framework called Language-Infused UNETR (LUNETR) was developed for 3D medical image segmentation by integrating visual and textual information through a hybrid CNN-Transformer architecture and cross-attention. The multimodal pancreatic dataset comprised CT images from 80 patients with pancreatic cancer at Zhuzhou Central Hospital and 55 CT image groups from the public Medical Segmentation Decathlon (MSD) dataset, resulting in a total of 135 pancreatic cancer cases with corresponding pathology text data. Incorporation of the semantic encoder improved the Dice score for pancreatic tumor segmentation by 2.23 percentage points. A 70%/10%/20% split was

used for training, validation, and testing, respectively. The authors also noted the model's dependence on large-scale annotated datasets and the need for strategies to improve generalization (59).

In another study, Du et al. demonstrated a Multi-Scale Channel Attention Unet (MSCA-Unet) architecture. They replaced the first convolutional layer of each multi-scale convolutional block of both the encoder and decoder with multiscale convolution. MSCA-Unet achieved a Dice index of 68.03% on a private dataset (Task-01) without data pre-processing and 80.12% on the public Medical Segmentation Decathlon dataset (Task-02) with data pre-processing. The architecture was specifically designed to improve segmentation of small and irregular pancreatic tumors (60).

Furthermore, Mukherjee et al. designed machine learning models to detect the imaging signature of pancreatic ductal adenocarcinoma (PDAC) using first-order intensity and second-order texture features obtained from volumetrically segmented normal pancreas tissue. They demonstrated an AUC of 0.98 (CI, 0.94–0.98) by categorizing the pancreas as normal versus prediagnostic for PDAC. However, the study's reliability and generalizability are compromised by the retrospective design, potential selection bias, and the small and comparatively younger validation cohorts (61).

Recent developments show AI's potential for PC, including 3D medical image segmentation, small-tumor segmentation, and prediagnostic PDAC detection. However, further investigation with larger sample sizes and broader analysis is required.

#### Methodological Robustness of the Current Evidence

The methodological strength of the reviewed studies differed considerably, particularly with respect to cohort size and external validation. For example, the KRAS-

prediction model reported by Song et al. was evaluated on 36 clinical cases and relied primarily on 15-fold cross-validation (31), whereas the GRAPE model was developed using data from two centers and subsequently evaluated in an independent external cohort across 16 centers (38). Cross-center generalizability also remains challenging: although HCA-DAN was specifically designed to reduce domain-related performance degradation, its application requires access to target-domain images and center-specific adaptation (43). Similarly, the CT-based radiomics nomogram of Huang et al. included external validation, but the external cohort comprised only 28

patients (44). These differences in cohort composition and validation strategy should be considered when interpreting apparently high model-performance estimates across studies. These considerations are consistent with contemporary reporting guidance for medical-imaging AI, which emphasizes transparent reporting of study design, image acquisition, data partitioning, internal and external testing, performance evaluation, and reproducibility (62). A structured comparison of the reviewed studies is presented in Table 1, while the principal methodological and translational bottlenecks affecting reproducibility and clinical translation are summarized in Table 2.

**Table 1:** Characteristics, performance, and main limitations of selected studies evaluating artificial intelligence and radiomics in the imaging of gastrointestinal cancer.

| Author           | Cancer type | AI approach/ Model name                                 | Purpose                                                                             | Imaging modality | Population / number of centers                                                                                                                                                                                                                                                         | Validation / metric source / key performance                                                                                                                                                                                                                        | External validation (independent cohort; external centres; external patients; metric source)                                                                                                                                                                                                                                               | Key limitations                                                                                                                                                   |
|------------------|-------------|---------------------------------------------------------|-------------------------------------------------------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Yang et al.(33)  | CRC         | Self-supervised learning-based/ <i>CRCFound</i>         | Preoperative staging and other diagnostic and prognostic tasks in colorectal cancer | CT               | CT images of 6332 CRC patients from 2008 to 2019 at the Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China.                                                                                                                                                         | In the external test dataset, <i>CRCFound_CT</i> achieved average AUCs of 0.7623 for T stage, 0.827 for N stage, and 0.7051 for M stage.                                                                                                                            | Yes; 1 external centre (Ningbo Medical Center Lihuili Hospital); n=120. TNM-stage AUCs were reported from the external test set; other task metrics were internally evaluated.                                                                                                                                                             | Limited performance in identifying complex samples                                                                                                                |
| Zhao et al.(34)  | CRC         | Deep learning / <i>PEDRA-EFBO</i>                       | Enhancing feature extraction and survival prediction in CRC CTs                     | CT               | 801 patients from the Affiliated Hospital of Jiangnan University in Wuxi City/November, 2007, to February, 2014                                                                                                                                                                        | The best performing combined feature set achieved a C-index of 0.9364, a Brier score of 0.0224, an MCC of 0.9738, an AUC of 0.9946, a sensitivity of 0.9936 and a specificity of 0.9984 using ten-fold cross-validation                                             | No independent external cohort reported; external centres=0; external patients=0; all listed metrics were obtained by internal 10-fold cross-validation.                                                                                                                                                                                   | Dependence on proprietary datasets                                                                                                                                |
| Song et al.(31)  | CRC         | Deep learning                                           | Identification of KRAS gene mutation                                                | T2 MRI           | 382 images from 36 clinical cases                                                                                                                                                                                                                                                      | Accuracy: 89.95%; Sensitivity: 89.29%; Specificity: 90.53%; AUC: 95.73%; Dice coefficient: 88.11%                                                                                                                                                                   | No; external centres=0; external patients=0; reported metrics were obtained primarily by internal 15-fold cross-validation.                                                                                                                                                                                                                | Small dataset (36 clinical cases; 382 images); evaluation primarily by 15-fold cross-validation                                                                   |
| Jia et al.(35)   | CRC         | Radiomics models                                        | KRAS mutation prediction                                                            | CT, MRI, PET     | Systematic review of 29 studies; meta-analysis of 19 studies                                                                                                                                                                                                                           | Pooled sensitivity, specificity, and AUC were 0.80, 0.80, and 0.87 in training cohorts and 0.78, 0.84, and 0.86 in validation cohorts                                                                                                                               | Mixed/NR meta-analysis; validation cohorts were reported in 13 studies, but external-centre and external-patient totals were not harmonized. Pooled validation metrics should not be interpreted as one model's external test.                                                                                                             | High heterogeneity and low methodological quality of the included radiomics studies; further external validation using multicenter datasets is required           |
| Hu et al.(38)    | GC          | Deep learning/ <i>GRAPE</i>                             | GC patients identification                                                          | CT               | Development cohort: 3,470 patients with gastric cancer and 3,250 non-gastric-cancer controls from two centers; internal validation: 1,298 cases; independent external validation: 18,160 cases from 16 centers; real-world evaluation: 78,593 consecutive noncontrast CT examinations. | <i>GRAPE</i> achieved an AUC of 0.970 in the internal validation cohort and 0.927 in the independent external validation cohort. Sensitivity was approximately 50% for early gastric cancer and exceeded 90% for T3 and T4 tumors.                                  | Yes; 16 external centres; n=18,160; external AUC=0.927 and internal AUC=0.970. The 78,593-case real-world evaluation was a separate observational evaluation.                                                                                                                                                                              | Lower sensitivity for early gastric cancer, approximately 50%, compared with more than 90% for T3 and T4 tumors.                                                  |
| Yuan et al.(43)  | GC          | Anisotropic neural network+ Transformer/ <i>HCA-DAN</i> | 3D CT tumor segmentation with domain adaptation                                     | CT               | 211 CT imaging cases from gastric cancer patients across four Chinese medical centers, including 211 non-enhanced and 63 contrast-enhanced CT volumes acquired with multiple CT scanners.                                                                                              | The AsTr baseline achieved DSCs of 59.26%, 55.97%, 48.83%, and 67.28% in four in-center tasks and 56.42%, 55.94%, 46.54%, and 60.62% in four cross-center tasks. <i>HCA-DAN</i> improved the cross-center DSCs to 58.36%, 56.72%, 49.25%, and 62.20%, respectively. | Cross-centre/domain-adaptation testing, not fully independent external validation; four centres in total, with one target centre per scenario (approximately 39–74 cases depending on the target centre). Cross-centre metrics were target-domain metrics obtained with adaptation; in-centre metrics were internal five-fold evaluations. | Target-domain dependency, retraining for each new center requirement, limited deployment in unseen clinical settings                                              |
| Huang et al.(44) | GC          | Radiomics nomogram                                      | Preoperative prediction of the histological type of gastric cancer                  | CT               | Retrospective cohort of 171 gastric cancer patients from two institutions: 143 patients for training/internal validation, 28 patients for external validation                                                                                                                          | AUCs of 0.777 (training), 0.726 (internal validation), and 0.720 (external validation)                                                                                                                                                                              | Yes; 1 external centre; n=28; external AUC=0.720. Internal AUC=0.726 and training AUC=0.777.                                                                                                                                                                                                                                               | Limited sample size and external validation; radiomics analysis was performed on portal venous phase CT only, with limited assessment of other clinical outcomes. |

| Author               | Cancer type                                   | AI approach/ Model name                                                                                                        | Purpose                                                          | Imaging modality                                               | Population / number of centers                                                                                                                                                                                                                                                    | Validation / metric source / key performance                                                                                                                                                                                                         | External validation (independent cohort; external centres; external patients; metric source)                                                                                                                                                                                                                    | Key limitations                                                                                                                                  |
|----------------------|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| Xie et al.(49)       | Liver cancer (HCC, ICC, and liver metastases) | Adaptive Balanced Dice Loss (BD Loss) with greedy parameter averaging (GPA) deep learning framework                            | Multi-class liver tumor segmentation                             | CT                                                             | 591 patients with malignant liver tumors                                                                                                                                                                                                                                          | DPC of 0.819 (95% CI: 0.798–0.841), VOE of 0.324 (95% CI: 0.268–0.380) and ASD of 4.69 (95% CI: 3.10–6.27) on the test set.                                                                                                                          | No independent external cohort reported; external centres=0; external patients=NR; reported test-set metrics were obtained from the collected study dataset.                                                                                                                                                    | Limited generalizability due to the absence of benign lesions in the evaluated cohort                                                            |
| Singh et al.(50)     | Liver cancer                                  | Hybrid deep learning framework combining VGG-16, ResNet-50, Channel Attention, and Spatial Attention mechanisms/ <i>FasNet</i> | Automated liver tumor segmentation                               | CT                                                             | LITS17 dataset                                                                                                                                                                                                                                                                    | Feature extraction and spatial feature selection for liver tumor segmentation improvement, achieving a Dice coefficient of 0.8766 and Jaccard Index of 0.8487                                                                                        | No study-specific external validation reported; external centres/patients=NR. Reported metrics were from the LITS17 benchmark evaluation and should not be labelled as an independent external clinical cohort.                                                                                                 | Further improvement of segmentation accuracy and validation using other imaging modalities (e.g., MRI) are required before clinical application. |
| Wu et al.(51)        | Liver (organ segmentation)                    | U-shaped network with multi-scale attention mechanism/ <i>MSA-UNet</i>                                                         | Liver organ segmentation in CT images                            | CT                                                             | 3Dircadb01 and MICCAI-Sliver07 dataset                                                                                                                                                                                                                                            | Test-set performance: DSC 98.00 ± 0.38%, IoU 96.08 ± 0.74%, precision 97.17 ± 0.85%, and recall 98.85 ± 0.70% for liver organ segmentation; MSA-UNet outperformed U-Net, U-Net++, DeepLabv3+, and CE-Net across these evaluation metrics.            | No independent external cohort reported; external centres=0; external patients=NR. The public datasets contained approximately 40 volumes and were split 8:1:1; metrics were from an internal hold-out test set.                                                                                                | Small sample size, heavy architecture of the network                                                                                             |
| Hua et al.(52)       | HCC                                           | Radiomics-based biomarker prediction                                                                                           | Prediction of vessels encapsulating tumor clusters (VETC) status | CEMRI, CECT, CEUS, [18F]FDG PET/CT                             | Meta-analysis of 15 studies                                                                                                                                                                                                                                                       | CEMRI: sensitivity 0.84, AUC 0.87; CEUS: specificity 0.81                                                                                                                                                                                            | Mixed meta-analysis; 7 of 15 studies contributed external-validation cohorts, comprising 613 external-validation patients in total. The number of external centres was NR, and pooled estimates combined internal and external data.                                                                            | Predominantly Chinese studies and generally small sample sizes, limiting generalizability                                                        |
| Shi et al.(59)       | Pancreatic cancer                             | Multimodal CNN–Transformer with an auto-encoding language encoder, cross-attention, and MSAA module                            | Segmentation of 3D medical images                                | Abdominal CT images combined with corresponding pathology text | 80 pancreatic cancer patients from Zhuzhou Central Hospital and 55 CT image groups from the public MSD dataset, resulting in a multimodal pancreatic dataset of 135 cases; additional evaluation was performed on the MSD liver (131 CT images) and lung (63 CT images) datasets. | The semantic encoder improved the average Dice score for pancreatic tumor segmentation by 2.23 percentage points; average Dice-score improvements of 4.31 and 3.67 percentage points were reported for the MSD liver and lung datasets, respectively | No separate independent external clinical cohort reported; external centres/patients=NR. The 135-case composite dataset was split internally (70%/10%/20%); public MSD data were part of the composite evaluation.                                                                                              | Dependence on large-scale annotated training data and requirement for text input during prediction, limiting clinical applicability.             |
| Du et al.(60)        | Pancreatic cancer                             | Deep learning, U-Net with multi-scale convolution and channel-attention mechanisms in the encoder and decoder                  | Pancreatic tumor segmentation                                    | CT                                                             | Private Task-01: 55 patients from the Affiliated Hospital of Qingdao University; public Task-02: 281 CT images from the Medical Segmentation Decathlon (MSD)                                                                                                                      | Dice 68.03% on private Task-01 without preprocessing and 80.12% on public MSD Task-02 with preprocessing; for tumors <100 mm <sup>2</sup> , the average Dice index was >0.68.                                                                        | No confirmed independent external validation. Public MSD Task-02 (n=281) was evaluated separately from private Task-01 (n=55), but independence from model development is not clearly established; therefore these results should be labelled benchmark/task-specific rather than external clinical validation. | Generalizability remains uncertain and requires further validation                                                                               |
| Mukherjee et al.(61) | Pancreatic ductal adenocarcinoma (PDAC)       | Radiomics-based machine learning (ML) model                                                                                    | PDAC at the pre-diagnostic stage detection                       | CT                                                             | 155 pre-diagnostic PDAC patients and 265 controls; training n=292, test n=128; internal validation n=176; NIH validation n=80                                                                                                                                                     | SVM: AUC 0.98, sensitivity 95.5%, specificity 90.3%, F1-score 89.5%, and accuracy 92.2%; specificity was 92.6% in the independent internal validation dataset and 96.2% in the public NIH validation dataset                                         | Yes; 1 external data source/centre (NIH Clinical Center); n=80 subjects. External specificity=96.2%; the reported AUC, sensitivity, specificity, F1-score, and accuracy were from the internal test set (n=128).                                                                                                | Retrospective design, possible selection bias, and small, relatively younger validation cohorts limiting reliability and generalizability        |

#### Abbreviations and footnotes:

NR, not reported; NA, not applicable. Cross-centre/domain-adaptation testing was not counted as fully independent external validation when target-domain data were used for adaptation. For meta-analyses, reported counts refer to contributing studies or validation cohorts and do not represent one model's external test set.

AI, artificial intelligence; ASD, average symmetric surface distance; AUC, area under the curve; BD loss, balanced Dice loss; C-index, concordance index; CECT, contrast-enhanced computed tomography; CEMRI, contrast-enhanced magnetic resonance imaging; CEUS, contrast-enhanced ultrasound; CI, confidence interval; CNN, convolutional neural network; CRC, colorectal cancer; CT, computed tomography; DPC, Dice per case; DSC, Dice similarity coefficient; F1-score, harmonic mean of precision and recall; GC, gastric cancer; GPA, greedy parameter averaging; GRAPE, Gastric Cancer Risk Assessment Procedure with AI; HCA-DAN, Hierarchical Class-Aware Domain-Adaptive Network; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; IoU, intersection over union; KRAS, Kirsten rat sarcoma viral oncogene homolog; LITS17, Liver Tumor Segmentation Challenge 2017; LUNETR, Language-Infused UNETR; MCC, Matthews correlation coefficient; ML, machine learning; MRI, magnetic resonance imaging; MSI, microsatellite instability; NA, not applicable; NR, not reported; PC, pancreatic cancer; PDAC, pancreatic ductal adenocarcinoma; PET/CT, positron emission tomography/computed tomography; TNM, tumor–node–metastasis; VETC, vessels encapsulating tumor clusters.

**Table 2.** Key methodological and translational bottlenecks in AI and radiomics for gastrointestinal cancer imaging

| Bottleneck                                     | Evidence from recent literature                                                                                                                                                                                                                                                                                                                                       | Methodological/translational consideration                                                                                                                                                                                          | Ref. |
|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| Radiomic feature stability and reproducibility | Lack of standardization in convolutional filtering can compromise reproducibility. In the IBSI standardization study, eight filter types were standardized, and 458 of 486 filtered-image features were reproducible across nine teams, using a lower 95% confidence bound of intraclass correlation coefficient >0.75 as the reproducibility criterion.              | Standardized filter definitions, reference images/feature values, and compliance testing can reduce implementation-related variability in radiomic feature extraction.                                                              | (67) |
| Scanner/protocol harmonization (ComBat)        | Radiomic features were sensitive to changes in MRI acquisition parameters, including flip angle, spatial resolution, section thickness, and parallel imaging. Across sequences, 51.4% of features were robust without ComBat, compared with 95.1% after parametric ComBat and 90.9% after nonparametric ComBat. The effect of harmonization differed among sequences. | ComBat can reduce acquisition-related variability, but transformation parameters should be estimated for the relevant imaging sequence because harmonization effects are not identical across sequences.                            | (74) |
| Class imbalance                                | In a CT cohort of 591 patients with different malignant liver-tumor types, data imbalance hindered multi-class segmentation. Balanced Dice loss mitigated this effect; the proposed method achieved a DPC of 0.819 (95% CI, 0.798–0.841), compared with 0.768 (95% CI, 0.740–0.796) for the baseline.                                                                 | Imbalance-aware optimization can support more balanced learning when tumor categories are unevenly represented.                                                                                                                     | (49) |
| Model explainability (Grad-CAM/SHAP)           | In a 2026 scoping review of 371 cancer-imaging studies, post-hoc XAI predominated (305/371, 82.2%). Visualization-based methods, including Grad-CAM, and feature-relevance approaches, including SHAP, were the major categories; however, only 193/371 studies reported any form of XAI validation.                                                                  | Post-hoc explanations should be accompanied by explicit validation. Greater standardization, quantitative validation, and user-centered evaluation are needed before XAI can be reliably integrated into clinical decision support. | (70) |

**Abbreviations:** AI, artificial intelligence; CI, confidence interval; CT, computed tomography; DPC, Dice per case; Grad-CAM, gradient-weighted class activation mapping; IBSI, Image Biomarker Standardization Initiative; MRI, magnetic resonance imaging; SHAP, Shapley additive explanations; XAI, explainable artificial intelligence.

## 2.5. Multimodal Data Integration and Clinical Utility

Clinical decision-making does not rely on imaging alone but integrates imaging findings with clinical, laboratory, and molecular information. Multimodal AI can combine heterogeneous inputs through early, joint/intermediate, or late fusion strategies; these approaches differ in whether information is integrated before, during, or after modality-specific learning (63). Cross-attention provides a more interactive approach to multimodal integration; in a 2024 study, a cross-attention architecture jointly incorporating CT imaging, gene-expression data, and clinical variables improved survival modeling in patients with non-small cell lung cancer (64). Evidence from gastrointestinal cancer also supports the added value of multimodal modeling: a recent colorectal-cancer study combined CT radiomics with clinical variables, including CEA and CA19-9, and showed that the fusion model provided greater net clinical benefit on decision curve analysis than either the radiomics or clinical model alone (65). Similarly, the gastric-cancer nomogram discussed in this review combined CT radiomic features with CA125 and demonstrated clinical net benefit on decision curve analysis (44). Therefore, the incremental value of radiomics should be assessed against clinical-only models using calibration and decision curve analysis, rather than inferred from discrimination metrics such as AUC alone.

## 2.6. Generalizability, Domain Shift, and Image Harmonization

Variability in image acquisition and preprocessing can affect radiomic feature reproducibility, making standardized preprocessing and transparent reporting important for multicenter analyses (66). Intensity normalization and discretization should be considered where appropriate, while voxel-size resampling is particularly

important when datasets contain heterogeneous voxel spacing or image resolution (66). For handcrafted radiomics, IBSI compliance should be verified against standardized reference images and feature values, as standardized feature extraction improves reproducibility across software implementations (67). Deep learning-based domain adaptation can reduce cross-center distribution shifts; in gastric CT, HCA-DAN improved cross-center segmentation, although its dependence on target-domain images may limit deployment in previously unseen centers (43). ComBat harmonization has also been shown to reduce manufacturer- and field-strength-related differences in radiomic features derived from multicenter abdominal MRI data (68). At the image level, CycleGAN-based harmonization improved inter-site radiomic-feature reproducibility and predictive performance in multicenter oncologic PET datasets (69).

## 2.7. Limitations

AI and radiomics have shown promising performance across multiple applications in gastrointestinal cancer imaging. However, there were some limitations in the reviewed studies. Most were retrospective, and based on single-center or geographically limited data sets, with limited generalizability of their findings. Validation was often limited or absent and a number of models were evaluated in relatively small or selected patient cohorts. Furthermore, the substantial variability in imaging protocols, AI architectures, clinical applications and performance evaluation metrics makes it difficult to directly compare the studies. Moreover, some studies reported lower model performance in challenging cases, early-stage disease, or small lesions, while certain approaches required extensive preprocessing, specialized datasets, or center-specific adaptation. Model interpretability remains an additional challenge for clinical translation. Post-hoc

approaches such as Grad-CAM and SHAP are widely used in cancer imaging, but their explanations are not consistently validated; therefore, visual or feature-attribution outputs should not automatically be interpreted as evidence of clinically meaningful model reasoning (70). Across the GI-cancer imaging literature reviewed here, no AI- or radiomics-based model has yet demonstrated improvement in a hard clinical endpoint, such as overall survival, disease-free survival, cancer-specific mortality, progression, treatment-related morbidity, or quality of life, in a prospective controlled trial. Therefore, diagnostic, segmentation, biomarker-prediction, prognostic, and decision-curve performance should not be interpreted as demonstrated patient benefit. Regulatory clearance or authorization alone does not establish clinical effectiveness or improved patient outcomes in routine care (71-73). Clinical translation also requires compliance with the relevant regulatory pathway, including 510(k) or De Novo authorization in the United States and conformity assessment under the European Medical Device Regulation for AI-based medical software in Europe (72). Taken together, these limitations emphasize that model performance should be interpreted alongside the strength of the underlying study design and validation strategy. Prospective multicenter evaluation, independent external testing across heterogeneous clinical and imaging environments, standardized and reproducible imaging workflows, and transparent reporting are required before AI and radiomics systems can be reliably incorporated into routine gastrointestinal cancer imaging.

### 3. Conclusion

AI and radiomics have shown potential across a range of applications in the imaging of colorectal, gastric, liver, and pancreatic cancers. Based on the studies reviewed, AI-

based methods showed potential in preoperative staging, tumor detection, automated segmentation, prediction of molecular biomarkers, histological classification, and prognostic assessment. Segmentation models showed variable performance across liver and pancreatic imaging studies, with higher Dice-based results reported in selected datasets. Radiomics and deep learning approaches showed potential for KRAS prediction, VETC assessment, gastric-cancer detection, and prediagnostic detection of pancreatic ductal adenocarcinoma. However, high performance in selected or internally validated datasets does not guarantee clinical application. Current evidence is limited by heterogeneity in study design, variation in validation strategies, limited diversity of datasets and lack of thorough evaluation in real-world settings. Therefore, AI and radiomics should be seen as decision support tools that complement but do not replace radiologists and multidisciplinary clinical assessment. The success of these systems in the future will depend on multicenter validation, reproducible data processing workflows, standardized reporting, improved model interpretability and evidence that such systems provide tangible benefits in terms of diagnostic efficiency, clinical decision making and patient outcomes.

### Abbreviation

AI: artificial intelligence  
 AJCC: American Joint Committee on Cancer  
 AUC: area under the curve  
 CAD: computer-aided detection  
 CA125: cancer antigen 125  
 CEMRI: contrast-enhanced magnetic resonance imaging  
 CEUS: contrast-enhanced ultrasound  
 CI: confidence interval  
 CNN: convolutional neural network  
 CRC: colorectal cancer  
 CT: computed tomography  
 DL: deep learning

DSC: Dice similarity coefficient  
 EUS: endoscopic ultrasonography  
 GC: gastric cancer  
 GI: gastrointestinal  
 HCC: hepatocellular carcinoma  
 ICC: intrahepatic cholangiocarcinoma  
 KRAS: Kirsten rat sarcoma viral oncogene homolog  
 LI-RADS: Liver Imaging Reporting and Data System  
 MRI: magnetic resonance imaging  
 MSI: microsatellite instability  
 PC: pancreatic cancer  
 PCR: polymerase chain reaction  
 PDAC: pancreatic ductal adenocarcinoma  
 PET: positron emission tomography  
 PET/CT: positron emission tomography/computed tomography  
 RO: microscopically margin-negative resection  
 TNM: tumor–node–metastasis  
 VETC: vessels encapsulating tumor clusters.

### Declarations

**Funding:** This work was supported by the National Institute for Medical Research Development (Grant No. 4002286 to Amir Avan).

**Clinical Trial Registration:** Not applicable. This article is a review and does not report the results of a clinical trial.

**Data availability:** Not applicable.

**Conflicts of Interest:** The authors have no relevant financial or non-financial interests to disclose.

**Ethics Approval:** Not applicable.

**Consent to Participate:** Not applicable.

**Consent for Publication:** Not applicable.

**Author Contributions Statement:** N.H., M.A., and R.R. contributed to Writing – Original Draft and Writing – Review & Editing. N.H., M.A., and A.Y.K. were responsible for the conceptualization, literature search, and drafting of the manuscript. I.S.G. contributed to data

curation and critical revision of the manuscript. A.G. was involved in methodological design and provided important intellectual content. A.A. and H.N. supervised the study, contributed to the study design, and performed critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

### Declaration of AI-assisted technologies:

During the preparation of this manuscript, ChatGPT (OpenAI) was used solely to improve grammar, language quality, clarity, readability, and section organization. All scientific content, numerical data, references, interpretations, and conclusions were independently reviewed and verified by the authors. The authors take full responsibility for the scientific integrity and final content of the manuscript.

### References

1. Organization WH. ICD-10: international statistical classification of diseases and related health problems: tenth revision: World Health Organization; 2004.
2. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, et al. Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology*. 2020;159(1):335–49.e15. <https://doi.org/10.1053/j.gastro.2020.02.068>
3. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, 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(3):229–63. <https://doi.org/10.3322/caac.21834>
4. Van Cutsem E, Nordlinger B, Cervantes A, Group EGW. Advanced colorectal cancer: ESMO Clinical Practice Guidelines for treatment. *Ann Oncol*. 2010;21 Suppl 5:v93–7. <https://doi.org/10.1093/annonc/mdq222>
5. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin*. 2017;67(3):177–93. <https://doi.org/10.3322/caac.21395>
6. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst*. 2004;96(19):1420–5. <https://doi.org/10.1093/jnci/djh275>

7. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209–49. <https://doi.org/10.3322/caac.21660>
8. Arnold M, Park JY, Camargo MC, Lunet N, Forman D, Soerjomataram I. Is gastric cancer becoming a rare disease? A global assessment of predicted incidence trends to 2035. *Gut*. 2020;69(5):823–9. <https://doi.org/10.1136/gutjnl-2019-320234>
9. Lin JL, Lin JX, Lin GT, Huang CM, Zheng CH, Xie JW, et al. Global incidence and mortality trends of gastric cancer and predicted mortality of gastric cancer by 2035. *BMC Public Health*. 2024;24(1):1763. <https://doi.org/10.1186/s12889-024-19104-6>
10. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. *Cancer Control*. 2017;24(3):1073274817729245. <https://doi.org/10.1177/1073274817729245>
11. Delghandi S, Raoufinia R, Shahtahmasbi S, Meshkat Z, Gouklani H, Gholoobi A. An overview of occult hepatitis B infection (OBI) with emphasis on HBV vaccination. *Heliyon*. 2024;10(17):e37097. <https://doi.org/10.1016/j.heliyon.2024.e37097>
12. Raoufinia R, Arabnezhad A, Keyhanvar N, Abdyazdani N, Saburi E, Naseri N, et al. Leveraging stem cells to combat hepatitis: a comprehensive review of recent studies. *Mol Biol Rep*. 2024;51(1):459. <https://doi.org/10.1007/s11033-024-09391-y>
13. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology*. 2018;68(2):723–50. <https://doi.org/10.1002/hep.29913>
14. Farges O, Fuks D, Le Treut YP, Azoulay D, Laurent A, Bachellier P, et al. AJCC 7th edition of TNM staging accurately discriminates outcomes of patients with resectable intrahepatic cholangiocarcinoma: By the AFC-IHCC-2009 study group. *Cancer*. 2011;117(10):2170–7. <https://doi.org/10.1002/cncr.25712>
15. Ayuso J-R, Balyaniskova S, Beets-Tan RGH, Blazic I, Blomqvist L, Caruso D, et al. MRI to guide clinical management of rectal cancer: updated consensus recommendations from the European Society of Gastrointestinal and Abdominal Radiology (ESGAR)—PART I primary staging. *European radiology*. 2026;36(6):4592–607. <https://doi.org/10.1007/s00330-025-12274-w>
16. Litjens G, Kooi T, Beijndorf BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>
17. Raoufinia R, Alimoradi M, Karimoddini M, Najafizadeh F, Taromi N, Bijari R, et al. Advancements in Personalized Medicine for Leukemia: Integrating Genetic, Transcriptomic, and Artificial Intelligence Insights. *Curr Pharm Des*. 2025. <https://doi.org/10.2174/0113816128402252251005215721>
18. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563–77. <https://doi.org/10.1148/radiol.2015151169>
19. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. *J Nucl Med*. 2020;61(4):488–95. <https://doi.org/10.2967/jnumed.118.222893>
20. Capello Ingold G, Martins da Fonseca J, Kolenda Złoić S, Verdan Moreira S, Kago Marole K, Finnegan E, et al. Preoperative radiomics models using CT and MRI for microsatellite instability in colorectal cancer: a systematic review and meta-analysis. *Abdom Radiol (NY)*. 2025;50(11):5148–61. <https://doi.org/10.1007/s00261-025-04981-1>
21. Jiang Y, Yuan H, Li Z, Ji X, Shen Q, Tuo J, et al. Global pattern and trends of colorectal cancer survival: a systematic review of population-based registration data. *Cancer Biology & Medicine*. 2022;19(2):175–86. <https://doi.org/10.20892/j.issn.2095-3941.2020.0634>
22. Doubeni CA, Corley DA, Quinn VP, Jensen CD, Zauber AG, Goodman M, et al. Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. *Gut*. 2018;67(2):291–8. <https://doi.org/10.1136/gutjnl-2016-312712>
23. Al Zaabi A, Al Maashri A, Bourdouce H, S AA-B. Emerging Endorobotic and AI Technologies in Colorectal Cancer Screening: A Review of Design, Validation, and Translational Pathways. *Diagnostics (Basel)*. 2026;16(3):421. <https://doi.org/10.3390/diagnostics16030421>
24. Atkin W, Dadswell E, Wooldrage K, Kralj-Hans I, von Wagner C, Edwards R, et al. SIGGAR investigators. Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. *Lancet*. 2013;381(9873):1194–202. [https://doi.org/10.1016/S0140-6736\(12\)62186-2](https://doi.org/10.1016/S0140-6736(12)62186-2)
25. Graser A, Stieber P, Nagel D, Schafer C, Horst D, Becker CR, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma

- in an average risk population. *Gut*. 2009;58(2):241–8.  
<https://doi.org/10.1136/gut.2008.156448>
26. Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, et al. Accuracy of CT colonography for detection of large adenomas and cancers. *N Engl J Med*. 2008;359(12):1207–17.  
<https://doi.org/10.1056/NEJMoa0800996>
  27. Kim DH, Pickhardt PJ, Taylor AJ, Leung WK, Winter TC, Hinshaw JL, et al. CT colonography versus colonoscopy for the detection of advanced neoplasia. *N Engl J Med*. 2007;357(14):1403–12.  
<https://doi.org/10.1056/NEJMoa070543>
  28. Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med*. 2003;349(23):2191–200.  
<https://doi.org/10.1056/NEJMoa031618>
  29. Nappi J, Yoshida H. Fully automated three-dimensional detection of polyps in fecal-tagging CT colonography. *Acad Radiol*. 2007;14(3):287–300.  
<https://doi.org/10.1016/j.acra.2006.11.007>
  30. Yoshida H, Dachman AH. Computer-aided diagnosis for CT colonography. *Semin Ultrasound CT MR*. 2004;25(5):419–31.  
<https://doi.org/10.1053/j.sult.2004.07.002>
  31. Song K, Zhao Z, Ma Y, Wang J, Wu W, Qiang Y, et al. A multitask dual-stream attention network for the identification of KRAS mutation in colorectal cancer. *Med Phys*. 2022;49(1):254–70.  
<https://doi.org/10.1002/mp.15361>
  32. Gheybi E, Hosseinzadeh P, Raoufinia R, Rostami M, Mahmoudi A, Soukhtanloo M. Contribution of Different Cell Signaling Pathways in Tumorigenesis. *Iranian Journal of Blood and Cancer*. 2026;18(2):37–46
  33. Yang J, Cai D, Liu J, Zhuang Z, Zhao Y, Wang FA, et al. CRCFound: A Colorectal Cancer CT Image Foundation Model Based on Self-Supervised Learning. *Adv Sci (Weinh)*. 2025;12(41):e07339.  
<https://doi.org/10.1002/advs.202407339>
  34. Zhao Z, Wang H, Wu D, Zhu Q, Tan X, Hu S, et al. PEDRA-EFBO: colorectal cancer prognostication using deep learning with patch embeddings and dual residual attention. *Med Biol Eng Comput*. 2025;63(6):1627–47.  
<https://doi.org/10.1007/s11517-025-03292-3>
  35. Jia LL, Zhao JX, Zhao LP, Tian JH, Huang G. Current status and quality of radiomic studies for predicting KRAS mutations in colorectal cancer patients: A systematic review and meta-analysis. *Eur J Radiol*. 2023;158:110640.  
<https://doi.org/10.1016/j.ejrad.2022.110640>
  36. Zhang G, Guo Y, Yang Y, Li T, Gong B, Lou J, et al. Radiogenomic landscape of the hallmarks of cancer. *Biomark Res*. 2026;14(1).  
<https://doi.org/10.1186/s40364-026-00920-4>
  37. Zhong J, Liu X, Lu J, Yang J, Zhang G, Mao S, et al. Overlooked and underpowered: a meta-research addressing sample size in radiomics prediction models for binary outcomes. *European radiology*. 2025;35(3):1146–56.  
<https://doi.org/10.1007/s00330-024-11331-0>
  38. Hu C, Xia Y, Zheng Z, Cao M, Zheng G, Chen S, et al. AI-based large-scale screening of gastric cancer from noncontrast CT imaging. *Nat Med*. 2025;31(9):3011–9.  
<https://doi.org/10.1038/s41591-025-03785-6>
  39. Pickhardt PJ, Summers RM, Garrett JW, Krishnaraj A, Agarwal S, Dreyer KJ, et al. Opportunistic screening: radiology scientific expert panel. *Radiology*. 2023;307(5):e222044.  
<https://doi.org/10.1148/radiol.222044>
  40. Giandola T, Maino C, Marrapodi G, Ratti M, Ragusi M, Bigioger V, et al. Imaging in Gastric Cancer: Current Practice and Future Perspectives. *Diagnostics (Basel)*. 2023;13(7):1276.  
<https://doi.org/10.3390/diagnostics13071276>
  41. Wang Z, Mo TM, Tian L, Chen JQ. Gastrin-17 Combined with CEA, CA12-5 and CA19-9 Improves the Sensitivity for the Diagnosis of Gastric Cancer. *Int J Gen Med*. 2021;14:8087–95.  
<https://doi.org/10.2147/IJGM.S335985>
  42. Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. *Gut*. 2019;68(9):1576–87.  
<https://doi.org/10.1136/gutjnl-2018-317556>
  43. Yuan N, Zhang Y, Lv K, Liu Y, Yang A, Hu P, et al. HCA-DAN: hierarchical class-aware domain adaptive network for gastric tumor segmentation in 3D CT images. *Cancer Imaging*. 2024;24(1):63.  
<https://doi.org/10.1186/s40644-024-00711-w>
  44. Huang H, Xu F, Chen Q, Hu H, Qi F, Zhao J. The value of CT-based radiomics nomogram in differential diagnosis of different histological types of gastric cancer. *Phys Eng Sci Med*. 2022;45(4):1063–71.  
<https://doi.org/10.1007/s13246-022-01170-y>
  45. Grazzini G, Cozzi D, Flammia F, Grassi R, Agostini A, Belfiore MP, et al. Hepatic tumors: pitfall in diagnostic imaging. *Acta Biomed*. 2020;91(8-s):9–17.  
<https://doi.org/10.23750/abm.v91i8-S.9969>
  46. Song L, Wang H, Wang ZJ. Bridging the Gap Between 2D and 3D Contexts in CT Volume for Liver and Tumor Segmentation. *IEEE J Biomed Health Inform*. 2021;25(9):3450–9.  
<https://doi.org/10.1109/JBHI.2021.3075752>
  47. Di S, Zhao YQ, Liao M, Zhang F, Li X. TD-Net: A Hybrid End-to-End Network for Automatic Liver Tumor Segmentation From CT Images. *IEEE J*

- Biomed Health Inform. 2023;27(3):1163–72.  
<https://doi.org/10.1109/jbhi.2022.3181974>
48. Jiang J, Peng Y, Hou Q, Wang J. MDCF\_Net: A Multi-dimensional hybrid network for liver and tumor segmentation from CT. *Biocybernetics and Biomedical Engineering*. 2023;43(2):494–506.  
<https://doi.org/10.1016/j.bbe.2023.04.004>
  49. Xie J, Zhou J, Yang M, Xu L, Li T, Jia H, et al. Lesion segmentation method for multiple types of liver cancer based on balanced dice loss. *Med Phys*. 2025;52(5):3059–74.  
<https://doi.org/10.1002/mp.17624>
  50. Singh R, Gupta S, Almogren A, Rehman AU, Bharany S, Altameem A, et al. FasNet: a hybrid deep learning model with attention mechanisms and uncertainty estimation for liver tumor segmentation on LiTS17. *Sci Rep*. 2025;15(1):17697.  
<https://doi.org/10.1038/s41598-025-98427-9>
  51. Wu J, Zhou S, Zuo S, Chen Y, Sun W, Luo J, et al. U-Net combined with multi-scale attention mechanism for liver segmentation in CT images. *BMC Med Inform Decis Mak*. 2021;21(1):283.  
<https://doi.org/10.1186/s12911-021-01649-w>
  52. Hua X, Fu R, Yin Z. Radiomics-Based AI for the Diagnosis and Prognosis of Vessels Encapsulating Tumor Clusters in Hepatocellular Carcinoma: Systematic Review and Meta-Analysis. *J Med Internet Res*. 2026;28:e90931.  
<https://doi.org/10.2196/90931>
  53. Yu Y, Cao L, Shen B, Du M, Gu W, Gu C, et al. Deep Learning Radiopathomics Models Based on Contrast-enhanced MRI and Pathologic Imaging for Predicting Vessels Encapsulating Tumor Clusters and Prognosis in Hepatocellular Carcinoma. *Radiol Imaging Cancer*. 2025;7(2):e240213.  
<https://doi.org/10.1148/rycan.240213>
  54. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7–33.  
<https://doi.org/10.3322/caac.21708>
  55. Walter FM, Mills K, Mendonca SC, Abel GA, Basu B, Carroll N, et al. Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. *Lancet Gastroenterol Hepatol*. 2016;1(4):298–306.  
[https://doi.org/10.1016/S2468-1253\(16\)30079-6](https://doi.org/10.1016/S2468-1253(16)30079-6)
  56. Khalaf N, El-Serag HB, Abrams HR, Thrift AP. Burden of Pancreatic Cancer: From Epidemiology to Practice. *Clin Gastroenterol Hepatol*. 2021;19(5):876–84.  
<https://doi.org/10.1016/j.cgh.2020.02.054>
  57. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nat Rev Gastroenterol Hepatol*. 2021;18(7):493–502.  
<https://doi.org/10.1038/s41575-021-00457-x>
  58. Huang B, Huang H, Zhang S, Zhang D, Shi Q, Liu J, et al. Artificial intelligence in pancreatic cancer. *Theranostics*. 2022;12(16):6931–54.  
<https://doi.org/10.7150/thno.77949>
  59. Shi Z, Zhang R, Wei X, Yu C, Xie H, Hu Z, et al. LUNETR: Language-Infused UNETR for precise pancreatic tumor segmentation in 3D medical image. *Neural Netw*. 2025;187:107414.  
<https://doi.org/10.1016/j.neunet.2025.107414>
  60. Du Y, Zuo X, Liu S, Cheng D, Li J, Sun M, et al. Segmentation of pancreatic tumors based on multi-scale convolution and channel attention mechanism in the encoder-decoder scheme. *Med Phys*. 2023;50(12):7764–78.  
<https://doi.org/10.1002/mp.16561>
  61. Mukherjee S, Patra A, Khasawneh H, Korfiatis P, Rajamohan N, Suman G, et al. Radiomics-based Machine-learning Models Can Detect Pancreatic Cancer on Prediagnostic Computed Tomography Scans at a Substantial Lead Time Before Clinical Diagnosis. *Gastroenterology*. 2022;163(5):1435–46.e3.  
<https://doi.org/10.1053/j.gastro.2022.06.066>
  62. Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiol Artif Intell*. 2024;6(4):e240300.  
<https://doi.org/10.1148/ryai.240300>
  63. Azarfar G, Naimimohasses S, Rambhatla S, Komorowski M, Ferro D, Lewis PR, et al. Responsible adoption of multimodal artificial intelligence in health care: promises and challenges. *Lancet Digit Health*. 2025;7(12):100917.  
<https://doi.org/10.1016/j.landig.2025.100917>
  64. Verma S, Magazzu G, Eftekhari N, Lou T, Gilhespy A, Occhipinti A, et al. Cross-attention enables deep learning on limited omics-imaging-clinical data of 130 lung cancer patients. *Cell Rep Methods*. 2024;4(7):100817.  
<https://doi.org/10.1016/j.crmeth.2024.100817>
  65. Jing HH, Hao D, Liu XJ, Cui MJ, Xue KJ, Wang DS, et al. Development and validation of a radiopathomics model for predicting liver metastases of colorectal cancer. *European radiology*. 2025;35(6):3409–17.  
<https://doi.org/10.1007/s00330-024-11198-1>
  66. Santinha J, Pinto Dos Santos D, Laqua F, Visser JJ, Groot Lipman KBW, Dietzel M, et al. ESR Essentials: radiomics-practice recommendations by the European Society of Medical Imaging Informatics. *European radiology*. 2025;35(3):1122–32.  
<https://doi.org/10.1007/s00330-024-11093-9>
  67. Whybra P, Zwanenburg A, Andrearczyk V, Schaer R, Apte AP, Ayotte A, et al. The Image Biomarker Standardization Initiative: Standardized Convolutional Filters for Reproducible Radiomics

- and Enhanced Clinical Insights. *Radiology*. 2024;310(2):e231319.  
<https://doi.org/10.1148/radiol.231319>
68. Jia W, Li H, Ali R, Shanbhogue KP, Masch WR, Aslam A, et al. Investigation of ComBat Harmonization on Radiomic and Deep Features from Multi-Center Abdominal MRI Data. *J Imaging Inform Med*. 2025;38(2):1016–27.  
<https://doi.org/10.1007/s10278-024-01253-0>
69. Haberl D, Spielvogel CP, Jiang Z, Orlhac F, Iommi D, Carrió I, et al. Multicenter PET image harmonization using generative adversarial networks. *Eur J Nucl Med Mol Imaging*. 2024;51(9):2532–46.  
<https://doi.org/10.1007/s00259-024-06708-8>
70. Fotopoulos D, Ladakis I, Filos D, Moreno-Sánchez PA, van Gils M, Chouvarda I. Explainable AI in Cancer Imaging: Scoping Review of Methods, Modalities, and Clinical Integration. *J Med Internet Res*. 2026;28:e80645.  
<https://doi.org/10.2196/80645>
71. Sivakumar R, Lue B, Kundu S. FDA Approval of Artificial Intelligence and Machine Learning Devices in Radiology: A Systematic Review. *JAMA Netw Open*. 2025;8(11):e2542338.  
<https://doi.org/10.1001/jamanetworkopen.2025.42338>
72. Zhang S, Li Y, Liu W, Chu Q, Wang S, Li J, et al. A decade of review in global regulation and research of artificial intelligence medical devices (2015-2025). *Front Med (Lausanne)*. 2025;12:1630408.  
<https://doi.org/10.3389/fmed.2025.1630408>
73. Song J, Gao Y, Liu W, Sun X, Chen C, Wu IX. Effectiveness of Artificial Intelligence-Assisted Examination for Cancer Detection in Medical Imaging: A Systematic Review and Meta-Analysis. *J Am Coll Radiol*. 2026;23(4):586–98.  
<https://doi.org/10.1016/j.jacr.2025.11.003>
74. Priya S, Dhruva DD, Sorensen E, Aher PY, Narayanasamy S, Nagpal P, et al. ComBat Harmonization of Myocardial Radiomic Features Sensitive to Cardiac MRI Acquisition Parameters. *Radiol Cardiothorac Imaging*. 2023;5(4):e220312.  
<https://doi.org/10.1148/ryct.220312>